

1 From collocations to call-ocations: using linguistic methods to quantify animal
2 call combinations

3

4 Alexandra B. Bosshard^{1,2*}, Maël Leroux^{1,2}, Nicholas A. Lester^{1,2}, Balthasar Bickel^{1,2}, Sabine
5 Stoll^{1,2} & Simon W. Townsend^{1,2,3}

6

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

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

9 ³Department of Psychology, University of Warwick, UK

10

11 *Corresponding author:

12 A. Bosshard, Department of Comparative Language Science, University of Zurich,

13 Thurgauerstrasse 30, 8050 Zurich, Switzerland.

14 *E-mail Address:* alexandra.bosshard@uzh.ch

15

16

17 **Abstract:**

18

19 Emerging data in a range of non-human animal species have highlighted a latent ability to
20 combine certain pre-existing calls together into larger structures. Currently, however, there
21 exists no objective quantification of call combinations. This is problematic because animal
22 calls can co-occur with one another simply through chance alone. One common approach
23 applied in language sciences to identify recurrent word combinations is collocation analysis.
24 Through comparing the co-occurrence of two words with how each word combines with
25 other words within a corpus, collocation analysis can highlight above chance, two-word
26 combinations. Here, we demonstrate how this approach can also be applied to non-human
27 animal communication systems by implementing it on a pseudo dataset. We argue
28 collocation analysis represents a promising tool for identifying non-random,
29 communicatively relevant call combinations in animals.

30

31 Key-words: call combinations, collocation analysis, comparative approach, non-random
32 structure, syntax

33

34 **Introduction**

35 Over the last 20 years there has been growing interest into the combinatorial abilities of
36 animals, namely the propensity to sequence context-specific calls (i.e. meaning-bearing
37 units, see (Suzuki & Zuberbühler, 2019) into larger potentially meaningful structures (Arnold
38 & Zuberbühler, 2006; Collier et al., 2020; Engesser, Ridley, & Townsend, 2016; Ouattara,
39 Lemasson, & Zuberbühler, 2009; Suzuki, Wheatcroft, & Griesser, 2016). Combinatorality is
40 one mechanism that can increase the expressive potential of a finite vocal repertoire. It
41 therefore provides important comparative insights into the complexity of animal vocal
42 systems and the selective pressures such systems have been exposed to (Collier et al.,
43 2020). These data also hold great promise in furthering our understanding of the similarities
44 between animal communication and human language given that, for many years, it was
45 assumed that the systematic concatenation of meaning-bearing units (i.e., syntax) was a
46 phenomenon unique to language (Hurford, 2012). Emerging examples of syntactic-like
47 structure in non-human primates and non-primate animals suggests this particular
48 assumption was indeed premature (Arnold & Zuberbühler, 2006; Berthet et al., 2019; Coye,
49 Ouattara, Zuberbühler, & Lemasson, 2015; Coye, Zuberbühler, & Lemasson, 2016) and such
50 data even have the potential to further our understanding of the evolutionary progression
51 of our own communication system (Leroux & Townsend, 2020; Townsend, Engesser, Stoll,
52 Zuberbühler, & Bickel, 2018).

53

54 In light of the communicative and evolutionary insights that research on combinatoriality
55 can provide, it is surprising that, to date, no objective means of quantifying call
56 combinations has been proposed. This is problematic as animal calls may occur in rapid
57 succession through chance alone, representing mere read-outs of contextual shifts. A

58 method to capture greater-than-chance co-occurrence of calls is therefore central to reliably
59 detecting and identifying non-random (i.e., potentially communicatively relevant) animal
60 call combinations.

61

62 Similar methodological issues have been encountered in research on language learning and
63 use (Bartsch, 2004; Evert, 2008; Gablasova, Brezina, & McEnery, 2017; Gries, 2013). One
64 approach frequently implemented to identify combinations of words (mostly bigrams i.e.,
65 two-word/two-call structures) in large written and spoken corpora is collocation analysis
66 (for a review, see Gries, 2013). Collocation analyses can take several forms, but the core
67 commonality is that it contrasts the frequency with which specific words combine to
68 measure the relative exclusivity of their relationship within a corpus (Church, Gale, Hanks, &
69 Hindle, 1991; Gries & Stefanowitsch, 2004; Kennedy, 1991; Nesselhauf, 2005). In other
70 words, such analyses reveal whether particular word/call combinations are more common
71 than would be expected given an assumed random baseline (e.g., the uniform distribution,
72 in which each combination is equally likely). For example, in English “*drink*” collocates with
73 “*coffee*” and “*going*” collocates with “*to*” (to form the future tense or describe a motion
74 event). Thus, collocation analyses can be understood as statistical measures of the influence
75 that a lexical item has on its neighbours. It has since become a crucial tool in Corpus
76 Linguistics for analysing lexical items and grammatical features of natural language (e.g.,
77 Bartsch, 2004; Lehecka, 2015; Stefanowitsch & Gries, 2003; Xiao & Mcenery, 2006).

78

79 In this paper we propose that by considering animal vocal data in a similar way as to how
80 language data are treated (i.e., as a corpus) affords the unique opportunity to apply a
81 variety of analytical tools habitually implemented in language sciences to study similar

82 questions. Specifically, we demonstrate the application of collocation analyses to non-
83 human animal datasets as a way to empirically identify combinations of two calls,
84 henceforth termed *bigrams*, and the relative merits of doing so. We focus on two specific
85 forms of collocation analysis commonly implemented in language sciences: Multiple
86 Distinctive and Mutual Information Collocation Analyses (Gries, 2014).

87

88 **Collocation analyses - Multiple Distinctive and Mutual Information approaches:**

89 Multiple Distinctive Collocation Analysis (MDCA) is primarily used when investigating and
90 testing for the degree of attraction between meaning-bearing units that share semantic
91 similarities in grammatical constructions. Specifically, MDCA statistically contrasts all
92 possible bigram combinations to estimate whether a given bigram occurs at frequencies
93 higher or lower than what would be expected by chance (Gries & Stefanowitsch, 2004;
94 Hilpert, 2006). Furthermore, the output of MDCA also provides a superficial estimate of
95 bigram ordering, namely whether the combination is sensitive to the position of the calls
96 comprising it (e.g., is A-B as frequent as B-A). Importantly, MDCA is not constrained by the
97 usual sampling assumptions, making them suitable for skewed, non-random, and small
98 corpora of the kind we tend to have in animal communication (Gries, 2014; Gries &
99 Stefanowitsch, 2004; Hilpert, 2006).

100 One recurrent issue for the analysis of linguistic corpora is the fact that any corpus
101 represents an incomplete – or undersampled – representation of the target linguistic system
102 – i.e., some two-word combinations can be under-represented or even absent from a
103 corpus when their “true” probability of occurrence is higher (note that this also affects all
104 other probabilities in the sample, which will be artificially inflated). Such undersampling
105 leads to misleading estimates of the significance of certain bigrams in the corpus (Gries &

106 Stefanowitsch, 2004; Hilpert, 2006). Mutual Information Collocation Analyses (MICA)
107 however actually overestimates to low-frequency values (Church & Hanks, 1990; Evert,
108 2005), which can be an advantage in animal communication corpora, as low-frequency
109 pairings, a common feature of non-human vocal data sets, are not overlooked but flagged as
110 potential combinatorial candidates. MICA calculates the variability of co-occurring items
111 through computing information values via observed frequency divided by expected
112 frequency. MICA can therefore – to some degree – better account for low-frequency
113 bigrams, a common occurrence in small corpora and particularly in non-human vocal
114 datasets.

115 In the remainder of the paper, we apply both forms of collocation analyses using an existing
116 R script provided by Stefan Gries (2014, see supplementary) to a pseudo data set we built
117 for the “Yeti” - a mythical ape-like creature.

118

119 *Call combinations in Yetis: an example:*

120 We created an artificial set of call combinations produced by the Yeti (see Table 1). The data
121 set was generated with a range of distributions in mind: specifically call combinations with
122 i) high-frequencies, ii) low frequencies, and call types that appear in combination with a)
123 only one other call type or b) with many different call types. The repertoire and sample size
124 were simulated consistent with the known communicative repertoires of other primates
125 (Leroux et al., in press). Yetis have a call repertoire comprising 10 calls, ranging from tonal
126 whistles and twitters to noisy barks and coughs. Initial assessment of the data suggested
127 that some of the calls co-occur with others from the repertoire and some co-occurrences
128 are more frequent than others. We therefore expected the collocation analyses to reveal at

129 least some significant associations between calls. To test this prediction, we applied both
 130 the MDCA and MICA to our data set.

131

132 Table 1: Distribution of bigrams occurring in the Yeti vocal repertoire. Columns and rows show the first and second unit
 133 within a call combination respectively.

		First unit									
		Bark	Cough	Cry	Howl	Huff	Hum	Peep	Puff	Twitter	Whistle
Second unit	Bark	-	-	-	-	-	-	-	-	-	-
	Cough	-	-	-	-	-	-	-	-	-	-
	Cry	-	3	-	-	-	-	-	-	-	-
	Howl	-	-	11	-	-	4	71	-	12	1
	Huff	-	-	-	-	-	-	-	-	-	-
	Hum	-	-	2	-	-	-	26	-	4	52
	Peep	-	-	-	-	-	-	-	-	-	-
	Puff	-	-	1	-	16	-	-	-	-	-
	Twitter	-	-	-	-	-	-	-	-	-	-
	Whistle	-	-	4	-	-	48	19	-	3	-

134
 135

136 *Multiple Distinctive Collocation Analysis:*

137 Sixteen different potential bigrams were identified in the Yeti repertoire (see Table 1).

138 In a first step we applied a Multiple Distinctive Collocation Analysis where call dependencies

139 within these bigrams were calculated using an exact binomial test on each possible bigram

140 combination (Gries, 2014; Gries & Stefanowitsch, 2004; Hilpert, 2006). Specifically, by

141 applying a logarithmic transformation to p-values (used here explicitly since they reflect the

142 relative association of calls but whilst simultaneously accounting for sample size), it is then

143 possible to estimate whether a given bigram occurs at frequencies higher (positive “pbins”,

144 Table 2) or lower (negative “pbins”, Table 2) than what would be expected by chance (i.e.
145 the absolute value of pbin >3: $P < 0.001$, $* > 2$: $P < 0.01$, $* > 1.3$: $P < 0.05$). Since the aim here is to
146 identify potential candidates for meaningful call combinations, we will focus only on positive
147 values that highlight an attraction between two call types. For positive pbin values, the
148 higher the value for two calls, the greater their collocational strength.

149 The Multiple Distinctive Collocation Analysis applied here suggests a significant relative
150 attraction exists within six bigrams (see Table 2). The highest value was calculated for Huff-
151 Puff, followed by Hum-Whistle, Whistle-Hum, Peep-Howl, Cough-Cry and lastly Twitter-
152 Howl. Importantly, four of the six bigrams showed significant attraction between the two
153 comprising calls in one specific order only (Huff-Puff, Peep-Howl, Cough-Cry & Twitter-Howl,
154 Table 2). However, the combinations involving Hums and Whistles displayed significant
155 attraction no matter the linearisation (Hum-Whistle and Whistle-Hum), suggesting that,
156 either order did not matter for this particular combination, or that Whistle and Hum form
157 two significant, differently ordered bigrams.

158

159 Table 2: Multiple Distinctive Collocation analysis for the bigrams in the Yeti’s vocal data set. Columns and rows show the
160 first and second unit within a call combination respectively. Values are pbins and can be translated to p-values (abs(pbin)
161 $* > 3$: $P < 0.001$, $* > 2$: $P < 0.01$, $* > 1.3$: $P < 0.05$). Significant results are coloured in *green*.

		First unit									
		Bark	Cough	Cry	Howl	Huff	Hum	Peep	Puff	Twitter	Whistle
Second unit	Bark	-	-	-	-	-	-	-	-	-	-
	Cough	-	-	-	-	-	-	-	-	-	-
	Cry	-	5.9	-0.1	-	-0.1	-0.3	-0.7	-	-0.1	-0.3
	Howl	-	-0.5	1.2	-	-2.6	-4.8	8.8	-	1.4	-7.7
	Huff	-	-	-	-	-	-	-	-	-	-
	Hum	-	-0.4	-1.1	-	-2.2	-7.6	-1.6	-	-0.5	17.0
	Peep	-	-	-	-	-	-	-	-	-	-
	Puff	-	-0.1	-0.2	-	18.6	-1.5	-4.0	-	-0.5	-1.6
	Twitter	-	-	-	-	-	-	-	-	-	-
	Whistle	-	-0.4	-0.3	-	-1.9	17.3	-2.6	-	-0.6	-6.8

162

163

164 In light of the additional aforementioned advantages associated with an information-based

165 approach, we complemented the MDCA with a MICA.

166

167 *Mutual Information Collocation Analysis:*

168 Similarly to MDCA, MICA calculates the collocational strength of each specific call type with

169 every other call type it collocates with. To do so, the *joint observed frequency* of a specific

170 bigram is divided by its *joint expected frequency* and then logarithmically transformed.

171 Concretely, the number of times the calls actually appear in combination is divided by the

172 number of times the calls would appear in combination if every call was randomly

173 distributed throughout the dataset. Once more, the higher the collocation value, the

174 stronger the collocational strength between two units (again, we focus on positive values

175 that indicate an attraction only). As with MDCA, pbins represent the logarithmically

176 transformed p-values (i.e., the absolute value of pbin >3: P<0.001, *>2: P<0.01, *>1.3:

177 P<0.05).

178 Not accounting for any specific ordering of the structures in the received input, the Mutual
 179 Information Collocation Analysis demonstrated a significant relative attraction within two
 180 bigrams (Cough-Cry & Huff-Puff, see Table 3). MICA only highlighted bigrams with call types
 181 that appear exclusively in combination with their collocational partner and not many other
 182 call types, while not rendering significant values for combinations of call types that appear
 183 with more than one other call type in the corpus (see Table 1). It is also noteworthy that the
 184 MICA provides a strong value for a very low-frequency pairing, namely Cough-Cry, which can
 185 only be found three times in the recorded repertoire.

186

187 Table 3: Mutual Information Collocation analysis for the bigrams in the Yeti's vocal data set. Columns and rows show the
 188 first and second unit within a call combination respectively. Values are pbins and can be translated to p-values (pbin >3 :
 189 $P < 0.001$, $* > 2$: $P < 0.01$, $* > 1.3$: $P < 0.05$). Significant results are coloured in green.

190

	Bark	Cough	Cry	Howl	Huff	Hum	Peep	Puff	Twitter	Whistle
Bark										
Cough	-									
Cry	-	3.7								
Howl	-	-	0.6							
Huff	-	-	-	-						
Hum	-	-	-2.4	-	-					
Peep	-	-	-	0.8	-	-1.1				
Puff	-	-	-0.4	-	4.0	-	-			
Twitter	-	-	-	0.8	-	-1.2	-	-		
Whistle	-	-	-1.3	-	-	0.7	-1.5	-	-1.5	

191

192 Discussion

193 Here, we show that, when conceptualising animal vocal data in the same way as a language
 194 corpus, methods habitually implemented in corpus linguistics can be transferred reliably to

195 non-human communication systems to highlight promising call combinations. We argue that
196 this approach therefore represents a novel application of a more objective method to
197 quantify the combinatorial dynamics of animal communication systems. Specifically,
198 collocation analyses help disentangle “true”, or non-random, call combinations from
199 happenstance juxtapositions of single calls. This is critical when investigating potentially
200 meaningful structuring within animal vocal communication systems.

201

202 It is important to note that other systematic approaches to capture the sequential dynamics
203 of animal vocal sequences (e.g., song), have been applied, including different Markovian and
204 non-Markovian chain modelling (Kershenbaum et al., 2014; Sainburg, Theilman, Thielk, &
205 Gentner, 2019; Suzuki, Buck, & Tyack, 2006). Nevertheless, we argue that there are strong
206 advantages for using collocation analysis, since collocations are easily operationalisable
207 across systems and provide a convenient and more descriptive (as opposed to modelling-
208 based) account of the combinatorial dependencies between vocal units (Evert, 2005, 2008;
209 Firth, 1957).

210

211 Whilst here we implement these analyses only with a pseudo data set, we have also applied
212 MDCA and MICA in real-world settings, primarily when investigating the combinatorial
213 properties of wild chimpanzee (Leroux et al., in press) and captive marmoset (Bosshard,
214 2020) vocal data sets. In these two primate examples, collocation analyses also reliably
215 identified non-random combinations and, critically, these included ones we had already
216 anecdotally highlighted as promising candidates in addition to others that were previously
217 less clear to us during observational data collection.

218

219 Of particular relevance is the fact that the type of collocational analyses applied here were
220 sensitive to bigrams even when they occurred very infrequently in the data set. This is
221 because collocational analyses consider the exclusivity of the combinatorial relationship: if
222 calls combine extremely rarely, they will still be detected as long as their relationship
223 together is exclusive. Since considerable variation characterises the frequency of call
224 combinations in animal communication (e.g., alarm call combinations are less frequent than
225 social call combinations (Boesch & Crockford, 2005; Collier, Townsend, & Manser, 2017;
226 Leroux, Chandia, Bosshard, Zuberbühler, & Townsend, in prep.)), we can be confident that
227 collocational analyses will identify all relevant combinations, both common and rare.

228

229 Another advantage of such collocation analyses is that they allow an estimate of the
230 ordering of call combinations. Identifying variation in the temporal progression of calls is
231 necessary to design experiments probing the role of order on meaning: data which are key
232 to unpacking how similar animal call combinations and human language really are. Our
233 results from the provided Yeti dataset indicate that some, but not all, of the identified
234 combinations are characterised by ordering, again a finding that is replicated in our real-
235 world data sets (Bosshard, 2020; Leroux et al., in press). This preliminary identification of
236 call order therefore might serve as one possible additional filter when deciding which of the
237 combinations detected from an animal data set to follow-up from an experimental/playback
238 perspective.

239

240 In conclusion, we hope that the approach outlined here will be applied by other researchers
241 in the field of animal communication as a way to disambiguate random from non-random
242 combinatorial structures. Future work could also build on these initial approaches through

243 applying other, as yet unexplored, sequence-based modelling methods currently used in
244 language sciences to animal corpora (e.g., skip-gram modelling, see Guthrie, Allison, Liu,
245 Guthrie, & Wilks, 2006). Ultimately, implementing the same objective, standardised
246 methods such as that presented in this paper could allow researchers to make more
247 meaningful comparisons both within and across systems, for example, at the individual,
248 group, population or even species level.

249

250

251 Acknowledgments:

252 We thank members of the Comparative Communication and Cognition Group at the
253 University of Zurich, specifically Piera Filippi, Mélissa Berthet, Joseph Mine, Anna J.
254 Szmarsowska and Manuel Rüdisühli for discussions. This work was supported by the Swiss
255 National Science Foundation (PP00P3_163850) to S.W.T. and the NCCR Evolving Language
256 (Swiss National Science Foundation Agreement #51NF40_180888). We declare no conflict of
257 interest.

258

259 **References**

- 260 Arnold, K., & Zuberbühler, K. (2006). Semantic combinations in primate calls. *Nature*,
261 441(7091), 303–303.
- 262 Bartsch, S. (2004). *Structural and functional properties of collocations in English: A corpus*
263 *study of lexical and pragmatic constraints on lexical co-occurrence*. Gunter Narr
264 Verlag.
- 265 Berthet, M., Mesbahi, G., Pajot, A., Cäsar, C., Neumann, C., & Zuberbühler, K. (2019). Titi
266 monkeys combine alarm calls to create probabilistic meaning. *Science Advances*,
267 5(5), eaav3991.
- 268 Boesch, C., & Crockford, C. (2005). Call combinations in wild chimpanzees. *Behaviour*,
269 142(4), 397–421.
- 270 Bosshard, A. B. (2020). *Sequential dynamics in common marmoset vocal strings*. University
271 of Zurich.
- 272 Church, K., Gale, W., Hanks, P., & Hindle, D. (1991). Using statistics in lexical analysis. *Lexical*
273 *Acquisition: Exploiting on-Line Resources to Build a Lexicon*, 115, 164.
- 274 Church, K., & Hanks, P. (1990). Word association norms, mutual information, and
275 lexicography. *Computational Linguistics*, 16(1), 22–29.
- 276 Collier, K., Radford, A. N., Stoll, S., Watson, S. K., Manser, M. B., Bickel, B., & Townsend, S.
277 W. (2020). Dwarf mongoose alarm calls: Investigating a complex non-human animal
278 call. *Proceedings of the Royal Society B*, 287(1935), 20192514.
- 279 Collier, K., Townsend, S. W., & Manser, M. B. (2017). Call concatenation in wild meerkats.
280 *Animal Behaviour*, 134, 257–269.

- 281 Coye, C., Ouattara, K., Zuberbühler, K., & Lemasson, A. (2015). Suffixation influences
282 receivers' behaviour in non-human primates. *Proceedings of the Royal Society B:*
283 *Biological Sciences*, 282(1807), 20150265.
- 284 Coye, C., Zuberbühler, K., & Lemasson, A. (2016). Morphologically structured vocalizations in
285 female Diana monkeys. *Animal Behaviour*, 115, 97–105.
- 286 Engesser, S., Ridley, A. R., & Townsend, S. W. (2016). Meaningful call combinations and
287 compositional processing in the southern pied babbler. *Proceedings of the National*
288 *Academy of Sciences*, 113(21), 5976–5981.
- 289 Evert, S. (2005). *The statistics of word cooccurrences: Word pairs and collocations*.
- 290 Evert, S. (2008). Corpora and collocations. *Corpus Linguistics. An International Handbook*, 2,
291 1212–1248.
- 292 Firth, J. R. (1957). *Modes of meaning Papers in linguistics 1934-1951 (pp. 190-215)*. London,
293 England: Oxford University Press.
- 294 Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in corpus-based language
295 learning research: Identifying, comparing, and interpreting the evidence. *Language*
296 *Learning*, 67(S1), 155–179.
- 297 Gries, S. T. (2013). 50-something years of work on collocations: What is or should be next....
298 *International Journal of Corpus Linguistics*, 18(1), 137–166.
- 299 Gries, S. T. (2014). *Coll.analysis 3.5. A script for R to compute perform collostructional*
300 *analyses*.
- 301 Gries, S. T., & Stefanowitsch, A. (2004). Extending collostructional analysis: A corpus-based
302 perspective on alternations'. *International Journal of Corpus Linguistics*, 9(1), 97–129.
- 303 Guthrie, D., Allison, B., Liu, W., Guthrie, L., & Wilks, Y. (2006). A closer look at skip-gram
304 modelling. *LREC*, 6, 1222–1225. Citeseer.

- 305 Hilpert, M. (2006). *Distinctive collexeme analysis and diachrony*.
- 306 Hurford, J. R. (2012). *The origins of grammar: Language in the light of evolution II* (Vol. 2).
307 Oxford University Press.
- 308 Kennedy, G. (1991). *Between and through: The company they keep and the functions they*
309 *serve*. na.
- 310 Kershenbaum, A., Bowles, A. E., Freeberg, T. M., Jin, D. Z., Lameira, A. R., & Bohn, K. (2014).
311 Animal vocal sequences: Not the Markov chains we thought they were. *Proceedings*
312 *of the Royal Society B: Biological Sciences*, 281(1792), 20141370.
- 313 Lehecka, T. (2015). Collocation and colligation. In *Handbook of pragmatics online*.
314 Benjamins.
- 315 Leroux, M., Bosshard, A. B., Chandia, B., Manser, A., Zuberbühler, K., & Townsend, S. W. (in
316 press). Chimpanzees combine pant-hoots with food-calls into larger structures.
317 *Animal Behaviour*.
- 318 Leroux, M., Chandia, B., Bosshard, A. B., Zuberbühler, K., & Townsend, S. W. (in prep.). *Call*
319 *combinations in chimpanzees: A social tool?*
- 320 Leroux, M., & Townsend, S. W. (2020). Call combinations in great apes and the evolution of
321 syntax. *Animal Behavior and Cognition*, 7(2), 131–139.
- 322 Nesselhauf, N. (2005). *Collocations in a learner corpus* (Vol. 14). John Benjamins Publishing.
- 323 Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys concatenate
324 vocalizations into context-specific call sequences. *Proceedings of the National*
325 *Academy of Sciences*, 106(51), 22026–22031.
- 326 Sainburg, T., Theilman, B., Thielk, M., & Gentner, T. Q. (2019). Parallels in the sequential
327 organization of birdsong and human speech. *Nature Communications*, 10(1), 1–11.

- 328 Stefanowitsch, A., & Gries, S. T. (2003). Collostructions: Investigating the interaction of
329 words and constructions. *International Journal of Corpus Linguistics*, 8(2), 209–243.
- 330 Suzuki, R., Buck, J. R., & Tyack, P. L. (2006). Information entropy of humpback whale songs.
331 *The Journal of the Acoustical Society of America*, 119(3), 1849–1866.
- 332 Suzuki, T. N., Wheatcroft, D., & Griesser, M. (2016). Experimental evidence for
333 compositional syntax in bird calls. *Nature Communications*, 7(1), 1–7.
- 334 Suzuki, T. N., & Zuberbühler, K. (2019). Animal syntax. *Current Biology*, 29(14), R669–R671.
- 335 Townsend, S. W., Engesser, S., Stoll, S., Zuberbühler, K., & Bickel, B. (2018). Compositionality
336 in animals and humans. *PLoS Biology*, 16(8), e2006425.
- 337 Xiao, Richard, & Mcenery, Tony. (2006). Collocation, Semantic Prosody, and NearSynonymy:
338 A Cross-Linguistic Perspective. *Applied Linguistics*, 27(1), 103–129.
339 <https://doi.org/10.1093/applin/ami045>
- 340