Rank-frequency analysis of characters in Garhwali text: emergence of Zipf's law

Manoj Kumar Riyal[1,*], Nikhil Kumar Rajput[2], Vinod Prasad Khanduri[1] and Laxmi Rawat[1]

[1]College of Forestry, VCSG Uttarakhand University of Horticulture and Forestry, Ranichauri, Tehri Garhwal 249 199, India
[2]Ramanujan College, University of Delhi, New Delhi, India

Zipf's law is ubiquitous in a language system, which establishes a relation between rank and frequency of characters or words. In the present study, it is shown that the distribution of character frequencies for Garhwali language follows Zipf–Mandelbrot law. Garhwali language is an Indo-Aryan language, spoken in the Garhwal region of Uttarakhand, India (northwestern Himalayan belt of India). The present communication examines the rank–frequency distribution by generalization of Zipf–Mandelbrot law in Garhwali language having limited dictionary size. The study shows that the distribution of character frequencies of consonants (with matras), vowels (including vowels with consonants in shape of matras) and all characters (including vowels and consonants without matras) for continuous Garhwali corpus follows Zipf–Mandelbrot law.

Keywords: Garhwali, frequency, rank, Zipf's law, Zipf–Mandelbrot law.

FREQUENCY–rank analysis plays an important role in many subjects, viz. physics, biology, linguistics, social sciences, etc. Word and character counting is the most popular discipline in the domain of quantitative linguistics. Instead of counting the words, we however have analysed the number of occurrences of characters used in our corpus, which is useful for growth of distinct words, birth and death of words. This study is also helpful for the process governing the usages of new words as well as language teaching, grammatical studies and entropic analysis[1–4].

Word counting, i.e. number of occurrences of the same word in a given corpus of a language is important for scientific study of any language. It is helpful for measuring the frequency and the rank of occurrences of words in the corpus. The Zipf's law states that $f(x) \sim 1/x$, i.e. the frequency $f(x)$ of the xth most frequent word decays in a database[5]. The character frequencies have also been analysed and decaying exponentially in the Zipf's plot[6]. Zipf's law is applicable for words both in English and characters of Chinese, however, it does not fit for all ranks[7]. The distribution of character frequencies of languages of English, French, Spanish, Italian has been followed Zipf's law[5,8,9]. The rank frequency distribution has been examined for four Indian languages, i.e. two Indo-Aryan and two Dravidian languages[10]. In randomly generated text, it has been observed that the frequency of occurrence of words almost follows inverse power law function of its rank and the exponent is close to 1 (ref. 11). Statistical analysis of the Indus script using *n*-grams also followed Zipf–Mandelbrot distribution[12]. Zipf–Mandelbrot law [$\log f = a - b \log(r + c)$] is the modified form of Zipf's law, where a and b are consonants and b is known as Zipf's exponent. For $c = 0$ and $b = 1$, this reduces to Zipf's law, $f = a/r$.

Several attempts have been made to find out the appropriate distribution for rank–frequency analysis. Few of the widely used distributions are Zipf–Mandelbrot, Zipf–Alekseev, negative hyper geometric and Lerch distribution[13]. Moreover, Zipf's law can be tested using zeta function, zeta distribution or in its simplest form of a power function[14]. Different models for rank frequency analysis on the grapheme frequencies for Slavik languages were tested[15,16] and it was found that negative hyper geometric distribution was appropriate for its description. A family of new exponential functions has also been proposed and tested[17,18]. It would be of interest to see these distributions in the Garhwali text.

In Garhwali language there are many characters that have their own meaning and significance like other languages, e.g. Chinese, Japanese and Korean[8]. So far, no such study has been carried out for analysis of Garhwali language. Garhwali language has many regional dialects spoken in different places scattered over a vast area in Uttarakhand because of the large mountainous chain-like structure. Most people belonging to northwestern part of Indian Himalaya speak Garhwali while some people speak Hindi, both of which lie in the Indo-Aryan linguistic group. The script used for writing Garhwali language is Devanagari[19].

The present study is important in extrapolating the limit of large databases and the results will be helpful in the field of agriculture extension (for information to the farmers and common people in Garhwali language), engineering, linguistics and physics for further study. This study aims to analyse the relation between frequency and rank of characters of the Garhwali language from Garhwali continuous corpus. Characters with maximum frequency have minimum rank, i.e. the most frequent character has rank 1; the next most frequent, rank 2, and so on.

Continuous Garhwali corpus was collected from different articles in an e-magazine containing 79,906 words[20]. The total number of characters in the corpus is 3, 17, 224. Furthermore, the frequency and its corresponding rank was measured for each vowels {/a/(अ); /a:/(आ); /i/(इ); /i:/ (ई); /u/(उ); /u:/(ऊ); /e/(ए); /e:/(ऐ); /o/(ओ); /o:/(औ)} as well as consonants {/ka/(क); /kha/(ख); /ga/(ग); /gha/(घ); /cha/(च); /chha/(छ); /ja/(ज); /jha/(झ); /chna/(ञ); /Ta/(ट); /Tha/(ठ); /Da/(ड); /Tna/(ण); /Dha/(ढ); /ta/(त); /tha/(थ); /da/(द); /dha/(ध);

*For correspondence. (e-mail: manoj.riyal@gmail.com)

Table 1. Rank and frequency of all consonants with matras

Character	Number of occurrences	Rank (r)	Frequency (f)	Character	Number of occurrences	Rank (r)	Frequency (f)
/ra/(र)	10324	1	0.347446	/vo/(वो)	473	63	0.015918
/na/(न)	8972	2	0.301945	/chaa/(चा)	472	64.5	0.015885
/ka/(क)	7332	3	0.246752	/ghe/(गे)	472	64.5	0.015885
/pa/(प)	6743	4	0.226930	/pha/(फ)	411	66	0.013832
/la/(ल)	5433	5	0.182843	/ri/(रि)	410	68	0.013798
/sa/(स)	4533	6	0.152554	/re/(रे)	410	68	0.013798
/maa/मा)	4032	7	0.135694	/sha/(श)	410	68	0.013798
/ma/(म)	3877	8	0.130477	/ke/(के)	393	70	0.013226
/da/(द)	3640	9	0.122501	/chhaa/(छा)	392	71.5	0.013192
/ja/(ज)	3233	10	0.108804	/sii/(सी)	392	71.5	0.013192
/ba/(ब)	3145	11	0.105842	/bu/(बु)	391	74	0.013159
/kaa/(का)	3045	12	0.102477	/pee/(पै)	391	74	0.013159
/Tna/(ण)	2954	13	0.099414	/phi/(फि)	391	74	0.013159
/ta/(त)	2754	14	0.092684	/lo/(लो)	371	76	0.012486
/cha/(च)	2657	15	0.089419	/knaa/(ज्ञा)	304	77	0.010231
/Ta/(ट)	2566	16	0.086357	/ko/(को)	302	78.5	0.010164
/vaa/(वा)	2245	17	0.075554	/lii/(ली)	302	78.5	0.010164
/ya/(य)	2231	18	0.075082	/ji/(जि)	301	80	0.010130
/ni/(नि)	2113	19	0.071111	/dhi/(धि)	301	81.5	0.010130
/ga/(ग)	2021	20	0.068015	/bha/(भ)	301	81.5	0.010130
/raa/(रा)	1982	21	0.066703	/ju/(जु)	294	83.5	0.009894
/rii/(री)	1974	22	0.066433	/dha/(ध)	294	83.5	0.009894
/ki/(कि)	1856	23	0.062462	/Tnu/(णु)	293	85	0.009861
/bi/(बि)	1793	24	0.060342	/shi/(शि)	291	86	0.009793
/yaa/(य ा)	1655	25	0.055698	/Tii/(टी)	280	87.5	0.009423
/ha/(ह)	1533	26	0.051592	/puu/(पू)	280	87.5	0.009423
/se/(से)	1501	27	0.050515	/dhaa/(धा)	279	89.5	0.009390
/jii/(जी)	1342	28	0.045164	/shaa/(शा)	279	89.5	0.009390
/mi/(मि)	1311	29	0.044121	/le/(ले)	271	91	0.009120
/va/(व)	1191	30	0.040082	/kee/(कै)	224	92	0.007539
/baa/(बा)	1054	31	0.035471	/tu/(तु)	213	94.5	0.007168
/laa/(ला)	1003	32	0.033755	/pu/(पु)	213	94.5	0.007168
/vi/(वि)	993	33	0.033419	/lee/(लै)	213	94.5	0.007168
/naa/(ना)	888	34	0.029885	/gu/(गु)	213	94.5	0.007168
/jaa/(जा)	883	35.5	0.029717	/haa/(हा)	211	97.5	0.007101
/daa/(दा)	883	35.5	0.029717	/pi/(पि)	211	97.5	0.007101
/chha/(छ)	880	37	0.029616	/ru/(रु)	209	99	0.007034
/ku/(कु)	854	38	0.028741	/yu/(यु)	208	100.5	0.007000
/taa/(ता)	830	39	0.027933	/chhe/(छे)	208	100.5	0.007000
/Da/(ड)	793	40	0.026688	/vu/(वु)	207	102	0.006966
/kii/(की)	791	41	0.026620	/ve/(वे)	203	103	0.006832
/di/(दी)	780	42	0.026250	/ho/(हो)	201	106	0.006764
/saa/(सा)	714	43	0.024029	/jo/(जो)	201	106	0.006764
/to/(तो)	706	44	0.023760	/Dha/(ढ)	201	106	0.006764
/ni/(नी)	680	45	0.022885	/vee/(वै)	201	106	0.006764
/mii/(मी)	680	46	0.022885	/Dii/(डी)	201	106	0.006764
/hii/(ही)	673	47	0.022649	/Tni/(णि)	183	109.5	0.006159
/chhoo/(छौ)	666	48	0.022414	/bee/(बै)	183	109.5	0.006159
/di/(दि)	618	50	0.020798	/ruu/(रू)	181	111	0.006091
/kha/(ख)	618	50	0.020798	/kho/(खो)	179	112.5	0.006024
/Tnaa/(णा)	618	50	0.020798	/Daa/(डा)	179	112.5	0.006024
/Tno/(णो)	609	52	0.020495	/no/(नो)	171	114.5	0.005755
/ti/(ति)	590	53	0.019856	/mee/(मै)	171	114.5	0.005755
/de/(दे)	584	54	0.019654	/hi/(हि)	170	116.5	0.005721
/gaa/(गा)	572	55	0.019250	/hu/(हु)	170	116.5	0.005721
/du/(दु)	510	57	0.017164	/khaa/(खा)	169	119	0.005688
/bhaa/(भा)	510	57	0.017164	/Tha/(ठ)	169	119	0.005688
/tha/(थ)	510	57	0.017164	/nu/(नु)	169	119	0.005688
/ghuu/(घू)	499	59	0.016793	/ro/(रो)	153	121.5	0.005149
/muu/(मु)	493	60	0.016592	/Ti/(टि)	153	121.5	0.005149
/paa/(पा)	480	61.5	0.016154	/Tnii/(णी)	151	123.5	0.005082
/li/(लि)	480	61.5	0.016154	/duu/(दू)	151	123.5	0.005082

(*Contd*)

Table 1. (*Contd*)

Character	Number of occurrences	Rank (r)	Frequency (f)	Character	Number of occurrences	Rank (r)	Frequency (f)
/po/(पो)	148	125.5	0.004981	/kna/(ज्ञ)	53	185.5	0.001784
/ye/(ये)	148	125.5	0.004981	/khii/(खी)	51	190	0.001716
/yo/(यो)	143	127	0.004813	/gha/(घ)	51	190	0.001716
/huu/(हू)	142	128.5	0.004779	/ghii/(घी)	51	190	0.001716
/he/(हे)	142	128.5	0.004779	/che/(चे)	51	190	0.001716
/Di/(डि)	141	131.5	0.004745	/je/(जे)	51	190	0.001716
/dhii/(धी)	141	131.5	0.004745	/poo/(पौ)	46	193	0.001548
/ne/(ने)	141	131.5	0.004745	/phii/(फी)	45	194.5	0.001514
/nee/(नै)	141	131.5	0.004745	/muu/(मू)	45	194.5	0.001514
/boo/(बौ)	135	134	0.004543	/sho/(शो)	43	197	0.001447
/do/(दो)	134	135	0.004510	/khe/(खे)	43	197	0.001447
/bii/(बी)	133	136.5	0.004476	/khee/(खै)	43	197	0.001447
/chu/(चु)	133	136.5	0.004476	/ghu/(घु)	41	199.5	0.001380
/choo/(चौ)	132	138.5	0.004442	/chhu/(छु)	41	199.5	0.001380
/Te/(टे)	132	138.5	0.004442	/cho/(छो)	40	201.5	0.001346
/Thee/(ठै)	124	140	0.004173	/jhe/(झे)	40	201.5	0.001346
/shha/(ष)	123	142	0.004139	/Tho/(ठो)	33	204	0.001111
/su/(सु)	123	142	0.004139	/te/(ते)	33	204	0.001111
/khi/(खि)	123	142	0.004139	/too/(तौ)	33	204	0.001111
/khu/(खु)	121	144	0.004072	/thii/(थी)	31	207	0.001043
/go/(गो)	119	146	0.004005	/bhe/(भे)	31	207	0.001043
/Tee/(णे)	119	146	0.004005	/bhee/(भै)	31	207	0.001043
/noo/(नौ)	119	146	0.004005	/bho/(भो)	29	209	0.000976
/be/(बे)	117	148	0.003938	/moo/(मौ)	28	210	0.000942
/jee/(जै)	107	149	0.003601	/si/(सि)	27	212.5	0.000909
/mo/(मो)	107	150.5	0.003601	/see/(सै)	27	212.5	0.000909
/koo/(कौ)	107	150.5	0.003601	/so/(सो)	27	212.5	0.000909
/gee/(गै)	105	152.5	0.003534	/hee/(है)	27	212.5	0.000909
/phaa/(फा)	105	152.5	0.003534	/khoo/(खौ)	26	216	0.000875
/vii/(वी)	84	154	0.002827	/gi/(गि)	26	216	0.000875
/soo/(सौ)	83	155.5	0.002793	/chhi/(छि)	26	216	0.000875
/phoo/(फौ)	83	155.5	0.002793	/Tu/(टु)	22	218.5	0.000740
/bhoo/(भौ)	82	157	0.002760	/Du/(डु)	22	218.5	0.000740
/me/(मे)	81	158	0.002726	/Doo/(डौ)	21	221	0.000707
/yi/(यि)	79	160	0.002659	/Dho/(ढो)	21	221	0.000707
/yii/(यी)	79	160	0.002659	/tho/(थो)	21	221	0.000707
/hoo/(हौ)	79	160	0.002659	/phuu/(फू)	13	223.5	0.000438
/gii/(गी)	78	162.5	0.002625	/phe/(फे)	13	223.5	0.000438
/Tee/(टै)	78	162.5	0.002625	/yuu/(यू)	9	226.5	0.000303
/Thii/(ठी)	75	166	0.002524	/vuu/(वू)	9	226.5	0.000303
/tee/(तै)	75	166	0.002524	/suu/(सू)	9	226.5	0.000303
/thi/(थि)	75	166	0.002524	/goo/(गौ)	9	226.5	0.000303
/bo/(बो)	75	166	0.002524	/ghaa/(घा)	8	230	0.000269
/lu/(लु)	75	166	0.002524	/ghi/(घि)	8	230	0.000269
/kuu/(कू)	74	169	0.002490	/ghoo/(घौ)	8	230	0.000269
/chee/(चै)	69	171.5	0.002322	/joo/(जौ)	7	234	0.000236
/De/(डे)	69	171.5	0.002322	/jha/(झ)	7	234	0.000236
/nuu/(नू)	69	171.5	0.002322	/jhi/(झि)	7	234	0.000236
/bhii/(भी)	69	171.5	0.002322	/jhu/(झु)	7	234	0.000236
/luu/(लू)	68	174	0.002288	/jhuu/(झू)	7	234	0.000236
/shu/(शु)	60	176	0.002019	/Too/(टौ)	6	237.5	0.000202
/guu/(गू)	60	176	0.002019	/Dhi/(ढि)	6	237.5	0.000202
/chi/(चि)	60	176	0.002019	/dhe/(धे)	5	239	0.000168
/chii/(ची)	59	178	0.001986	/pe/(पे)	5	241	0.000168
/To/(टो)	58	179.5	0.001952	/phee/(फै)	5	241	0.000168
/Do/(डो)	58	179.5	0.001952	/buu/(बू)	5	241	0.000168
/dee/(दै)	57	182	0.001918	/bhi/(भि)	4	244.5	0.000135
/dhu/(धु)	57	182	0.001918	/shee/(शै)	4	244.5	0.000135
/pho/(फो)	57	182	0.001918	/shhaa/(षा)	4	244.5	0.000135
/ree/(रै)	53	185.5	0.001784	/shhi/(षि)	4	244.5	0.000135
/roo/(रौ)	53	185.5	0.001784	/khuu/(खू)	3	247	0.000101
/shii/(शी)	53	185.5	0.001784				

Table 2. Frequency and rank of all Garhwali vowels (with matras in consonants) used in corpus

Character	Number of occurrences	Rank (r)	Frequency (f)
/a/(अ)	83544	1	0.263359645
/a:/(आ)	33231	2	0.10475563
/i/(इ)	21711	3	0.068440597
/i:/(ई)	13221	4	0.041677174
/e/(ए)	11643	5	0.036702772
/u/(उ)	10648	6	0.033566187
/o/(ओ)	5434	7	0.017129851
/e:/(ऐ)	3334	8	0.010509924
/o:/(औ)	2583	9	0.008142511
/u:/(ऊ)	2335	10	0.007360729

Table 3. Frequency and rank of all Garhwali characters (vowels and consonants without matras) used in the corpus

Character	Number of occurrences	Rank (r)	Frequency (f)
/ra/(र)	11598	1	0.036560916
/na/(न)	9873	2	0.031123118
/ka/(क)	7622	3	0.024027186
/la/(ल)	6711	5	0.021155398
/pa/(प)	6588	4	0.020767659
/ma/(म)	5642	7	0.01778554
/sa/(स)	5387	6	0.016981691
/da/(द)	4322	8	0.013624442
/a/(अ)	4298	9	0.013548786
/ja/(ज)	4190	10	0.013208332
/ba/(ब)	3892	11	0.012268933
/Tna/(ण)	3791	12	0.011950546
/ta/(त)	3541	13	0.011162459
/cha/(च)	3325	14	0.010481552
/Ta/(ट)	3121	15	0.009838474
/ya/(य)	2871	16	0.009050387
/ga/(ग)	2612	17	0.008233929
/a:/(आ)	2251	18	0.007095932
/i/(इ)	2125	19	0.006698737
/ha/(ह)	1877	20	0.005916955
/va/(व)	1632	21	0.00514463
/e/(ए)	1490	22	0.004696996
/chha/(छ)	1288	23	0.004060222
/Da/(ड)	1155	24	0.00364096
/tha/(थ)	955	27	0.003010491
/kha/(ख)	879	25	0.002770913
/u/(उ)	744	26	0.002345346
/pha/(फ)	604	28	0.001904017
/sha/(श)	601	29	0.00189456
/bha/(भ)	549	30	0.001730638
/dha/(ध)	509	31	0.001604544
/Dha/(ढ)	508	32	0.001601392
/i:/(ई)	471	33	0.001484755
/Tha/(ठ)	432	34	0.001361814
/o/(ओ)	311	35	0.00098038
/shha/(ष)	261	36	0.000822762
/e:/(ऐ)	105	37	0.000330996
/u:/(ऊ)	101	38	0.000318387
/o:/(औ)	74	39	0.000233274
/kna/(ज्ञ)	63	40	0.000198598
/gha/(घ)	28	41	8.82657E-05
/jha/(झ)	19	42	5.98946E-05

/na/(न); /pa/(प); /pha/(फ); /ba/(ब); /bha/(भ); /ma/(म); /ya/(य); /ra/(र); /la/(ल); /va/(व); /sha/(श); /shha/(ष); /sa/(स); /ha/(ह); /kna/(ज्ञ)}.

The frequency f of any character can be defined as the ratio of the number of occurrences of the character to the total number of characters in the text, i.e. if in part i, N_i is the total number of characters or words and n_i the number of occurrences of a given character or word in that part, the ratio n_i/N_i gives the frequency f of appearance of the word or character of that part.

The rank–frequency curve has been plotted and observed that it follows the modified form of Zipf's law, i.e. Zipf–Mandelbrot distribution[12]. In more generalized form, if a large corpus is considered and ranks assigned to all words according to the order of decreasing frequencies of those words, then the frequency $f(x)$ of a word of rank x satisfies the relation $f(x) = ax^b$, where a and b are constants and b is the Zipf's exponent.

The number of occurrences of each character was measured and arranged in decreasing frequency. Frequency and their corresponding rank for each character were then calculated. Table 1 gives the frequency and their corresponding ranks of all consonants used in the study. Similarly, Table 2 gives the frequency and rank of all Garhwali vowels (including matras in consonants). The frequency and rank of all Garhwali characters (without matras) used in the corpus are given in Table 3. The frequency–rank plot (log–log scale) for consonants (with matras), the frequency–rank plot (log–log scale) of vowels used in corpus (including the vowels with consonants in shape of matras) and the frequency–rank plot (log–log scale) of all characters (including vowels and consonants without matras) are shown in Figures 1–3 respectively.

The data fitted well with the Zipf–Mandelbrot law, $\log f = a - b \log(c + r)$. Results show that the character frequency distribution follows an inverse power law with exponent, whereas the character frequency decays exponentially in the Zipf's plot. From the graph, it is clear that the characters used in the corpus follow Zipf–Mandelbrot

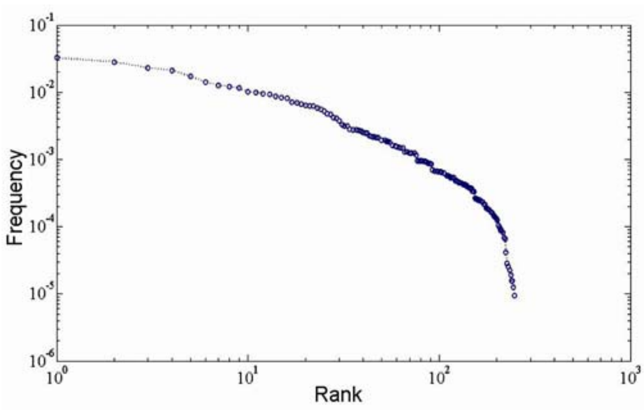

Figure 1. Frequency–rank plot (log–log scale) for consonants (with matras) used in the corpus with parameters $a = 0.4144$ and $b = -0.7151$.

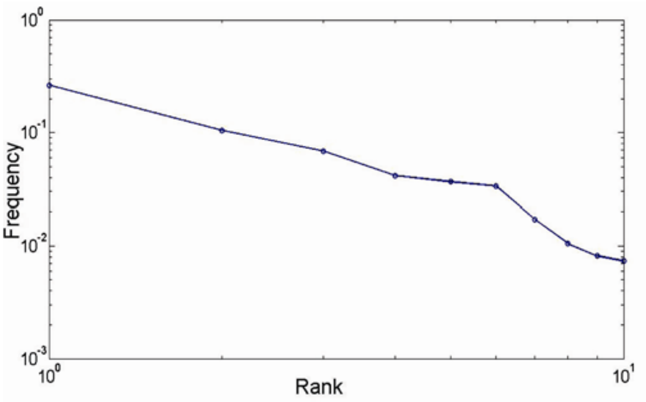

Figure 2. Frequency–rank plot (log–log scale) of vowels (including consonants) used in corpus with parameters $a = 0.2637$ and $b = -1.307$.

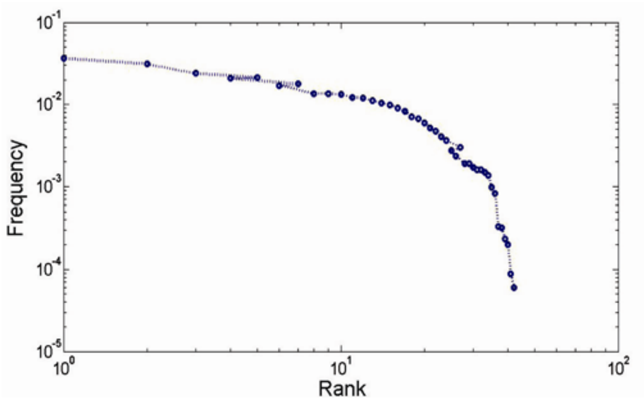

Figure 3. Frequency–rank plot (log–log scale) of all Garhwali characters (vowels and consonants without matras) used in corpus with parameters $a = 0.0.04346$ and $b = -0.6402$.

law (inverse power law function) for the distribution of character frequencies of consonants (with matras), vowels (including vowels with consonants in shape of matras) and all characters (including vowels and consonants without matras) respectively (Figures 1–3).

The graph between log frequency and log rank shows that the frequency of occurrences of characters in all the three cases follow Zipf–Mandelbrot law ($f(x) = ax^b$ or $\log f(x) = \log a - b \log x$) and character frequency decays almost exponentially in the Zipf's plot. The log–log plot of frequency and rank of consonants (with matras) with $a = 0.63505$ and $b = -0.6793$ (Figure 1), the log–log graph of frequency and rank of all vowels with parameter values, $a = 0.246$ and $b = -1.495$ (Figure 2) and the log–log plot of frequency and rank vowels and consonants (without matras) with $a = 0.0329$ and $b = -0.6411$ (Figure 3) are also shown. Since this study shows that the value of Zipf's exponent (b) is not equal to –1, the characters of Garhwali language in continuous corpus are in good agreement with the empirical rank–frequency distribution, i.e. Zipf–Mandelbrot distribution.

The results obtained in this study have been compared with those of Chinese language[21]. The frequency of characters in the Chinese, Japanese and Korean language follow the power law with exponent close to 1 (ref. 8). The distinct element grows with stable exponent if the frequency of the characters follows Zipf's law[22].

The results of this study give an overview of the frequency of character in Garhwali language which follows Zipf–Mandelbrot distribution. The present study gives a good estimation of the Zipf's exponent in a Garhwali text. The study on Garhwali language is meagre so far. Garhwali is one of the languages of India which is in the list of elimination identified by UNESCO[23]. This study is beneficial for agricultural extension, computer science, engineering and linguistics students for further research. In physics, this study would be helpful to augment stochastic model for vocabulary growth of Garhwali language. Rank–frequency analysis is also essential to understand the birth and death of words as well as the process governing the addition of new words which can be used for the development of stochastic modeling. In linguistics, the study is beneficial to know which character has the maximum occurrence in Garhwali language and will support the development of Garhwali sound system, i.e. to see the properties of characters in words initial, middle and final position. In engineering, the study will be helpful for the development of automatic speech recognition system of Garhwali language with the help of the basic properties, i.e. rank–frequency of Garhwali language, which will help the common people as well as the farmers and the natives of the area to obtain information in their own language.

1. Gerlach, M. and Altmann, E. G., Stochastic model for the vocabulary growth in natural languages. *Phys. Rev. X*, 2013, **3**; doi:10.1103/PhysRevX.3.021006.
2. Montemurro, M. A. and Zanette, D. H., Entropic analysis of the role of words in literary texts. *Adv. Complex Syst*., 2002, **5**(1), 7–17.
3. Pande, H. and Dhami, H. S., Mathematical modelling of occurrence of letters and word's initials in texts of Hindi language. *J. Theor. Ling.*, 2010.
4. Wyllys, R. E., Empirical and theoretical bases of Zipf's law. *Libr. Trends*, 1981, **30**(1), 53-64.
5. Zipf, G. K., *Human Behaviour and the Principal of Least Effort*, Wesley, Reading, 1949.
6. Lü, L., Zhang, Z. and Zhou, T., Deviation of Zipf's and Heaps' laws in human languages with limited dictionary sizes. *Sci. Rep.*, 2013, **3**; doi:10.1038/srep01082.
7. Ha, L. Q., Sicilia-Garcia, E. I., Ming, J. and Smith, F. J., Extension of Zipf's Law to word and character N-grams for English and Chinese. *Comput. Ling. Chinese Lang. Proc.*, 2003, **8**(1), 77–102.
8. Cancho, R. F. I. and Solé, R. V., Least effort and the origins of scaling in human language. *Proc. Natl. Acad. Sci. USA*, 2002, **100**, 788–791.
9. Kanter, I. and Kessler, D. A., Markov processes: linguistics and Zipf's law. *Phys. Rev. Lett.*, 1995, **74**, 4559–4562.
10. Jayaram, B. D. and Vidya, M. N., Zipf's law for Indian Languages. *J. Quant. Ling.*, 2008, **15**(4), 293–317.
11. Li, W., Random texts exhibit Zipfs-law-like word frequency distribution. *IEEE Trans. Inf. Theory*, 1992, **38**(6), 1842–1845.
12. Yadav, N., Joglekar, H., Rao, R. P. N., Vahia, M. N., Adhikari, R. and Mahadevan, I., Statistical analysis of the Indus script using *n*-grams. *PLoS ONE*, 2010, **5**(3), e9506.

13. Tuzzi, A., Popescu, I. and Altmann, G., Zipf's law Italian texts. *J. Quant. Ling.*, 2009, **16**(4), 354–367.
14. Zornig, P. and Altmann, G., Unified representation of Zipf's distributions. *Comput. Stat. Data Anal.*, 1995, **19**, 461–473.
15. Gryzbek, P. and Kelih, E., Towards a general model of grapheme frequencies for Slavic languages. In Proceeding of Third International Seminar on Computer Treatment of Slavic and East European Languages, Bratislva, Slovakia, 2005.
16. Gryzbek, P., Kelih, E. and Stadlober E., Slavic letter frequencies: A common discrete model and regular parameter behaviour? *Iss. Quant. Ling.*, 2009, 17–33.
17. Gunther, R., Levitin, L., Schapiro, B. and Wagner, P., Zipf's law and the effect of ranking on probability distributions. *Int. J. Theor. Phys.*, 1996, **35**(2).
18. Popescu, I., Cech, R. and Altmann, G., *The Lambda Structure of Texts*, Ram-Verlag, 2011, pp. 181.
19. http://en.wikipedia.org/wiki/Garhwali_language
20. http://e-agazineofuttarakhand.blogspot.in/2009/10/garhwali-kumaoni-himalayan-literature_6723.html
21. Wang, D. H., Li, M. H. and Di, Z. R., Ture reason for Zipf's law in language. *Physica A*, 2005, **358**, 545–550.
22. Lü, L., Zhang, Z.-K. and Zhou, T., Zipf's law lwads to Heaps' law: analysing their relation in finite-size systems. *PLoS ONE*, 2010, **5**, e14139.
23. http://www.language-archives.org/language/gbm#language_descriptions

Received 2 January 2015; revised accepted 30 September 2015

doi: 10.18520/cs/v110/i3/429-434

Life history fitness of giant ladybird predator (Coleoptera: Coccinellidae) of woolly aphids (Hemiptera: Aphididae) in varying prey densities from northeast India

Basant K. Agarwala* and Joydeb Majumder

Ecology and Biodiversity Laboratories, Department of Zoology, Tripura University, Suryamaninagar, Tripura 799 022, India

Quantity of natural prey available to insect predators varies in time and space and it influences the fitness of individual predators. This was examined for *Anisolemnia dilatata* (Fab.) which is a specialist predator of woolly aphids of bamboo plants and sugarcanes, and endemic to south Asia and Asia-Pacific regions. Results of a laboratory study using 45 larvae and 10 adult females showed that individual *A. dilatata* larvae performed best at an optimal density of 250 live aphids per larva per day, and adult females from these larvae matured faster and produced higher number of viable eggs than the larvae that grew and developed at sub-optimal prey density of 200 or less aphids. About 73–80% of the larvae that survived at the density of 50 or 100 or 150 aphids per day took significantly longer time to complete development and to reach the age at maturity. Such females were significantly smaller in size and produced fewer viable eggs. Results showed that larvae and adults of *A. dilatata* required high density of aphid prey to support its optimal life history fitness. Results hold promise in the application of this predator in the control of waxy aphid pest of cultivated bamboos and sugarcanes.

Keywords: Bamboo plants, giant ladybird predator, life history fitness, prey requirement, woolly aphids prey.

*For correspondence. (e-mail: bagarwala00@gmail.com)

ANISOLEMNIA dilatata (Fab.) (Coleoptera: Cocinellidae) is one of the two prevalent giant ladybird beetle predators of waxy aphids that occurs in tropical seasonal forests of South, South East and Far East Asia[1–3] which contain bamboos as one of the principal component of vegetation complex. The other species is *Synonycha grandis* Thunberg[4]. Giant ladybird beetles are unique among ladybird predators due to their large size and prey specialization of woolly aphids infesting bamboos and sugarcanes[5]. Imagos of *A. dilatata* measure about 3.5 times (mean ± SE = 137.43 ± 0.68 mg, $n = 45$) in comparison to the average size of aphidophagous *Coccinella septempunctata* L. (mean ± SE = 39.42 ± 1.22 mg, $n = 20$) that feeds on several aphid species in diverse habitats[6,7]. Out of 78 species of bamboos under 19 genera recorded from northeast India, *Bambusa balcooa* Roxb., *B. aurandinacea* (Retz.) Willd. and *B. tulda* Roxb. are widely cultivated and used for economic purposes in Tripura[8–10], a province in the south of northeast India. Wool producing horned aphid species, *Ceratovacuna silvestrii* (Takahashi), *C. indica* (Ghosh, Pal & Raychaudhury), *Paraoregma alexendari* (Takahashi) and *Pseudoregma bucktoni* (Takahashi), make dense colonies on young leaves and tender shoots of perennial bamboo species. Among the several natural enemies of woolly aphids that have been recorded on bamboo plants, larvae and adults of *A. dilatata* are found to be the dominant predators (Figure 1 *a* and *c*)[11].

Biology and ecology of smaller species of ladybird predators of aphids and coccids are well known with respect to their functional and numerical responses to different prey species[12,13], but no information exist on the life history fitness parameters of the giant ladybird species, *A. dilatata*. In a field study from Tripura, Majumder and Agarwala[11] showed that the incidence of *A. dilatata* on two bamboo species, *B. tulda* and *B. balcooa*, was restricted to high density phase of aphid population of *C. silvestrii*; thus populations of predator and prey co-existed for only 27 weeks out of 44 weeks incidence of the aphid prey in a year of study. Due to economic importance of bamboo plants as means of livelihood of large number of rural folk in parts of south Asia[8] and great prey-feeding potential of *A. dilatata*[4,5], a laboratory study