Bengali language and Graphical law

Anindya Kumar Biswas, Department of Physics;
North-Eastern Hill University, Mawkynroh-Umshing, Shillong-793022.
email:anindya@nehu.ac.in

January 22, 2020

Abstract

We study a bengali to bengali dictionary. We draw in the log scale, number of words starting with a letter vs rank of the letter, both normalised. We find that the graphs are closer to the curves of reduced magnetisation vs reduced temperature for various approximations of Ising model.

I Introduction

"Sobdoi Brahma(Word is Supreme)": Quote unknown

Our world is beaming with languages. Some are spoken by many. Some are spoken by few. English is a language known to large part of the world. Official language in most part of the world is English. Dictionaries from English to a natural language is of common availability. A Dictionary is where words beginning with letters, are recorded in their relative abundances. Few letters are heavily used. Few are scantily done so. Number of letters vary from a language to another language.

Magnetic field is omnipresent in our habitat. From down it is earth's magnetic field. Above on our head strating from atmospheric current and associated magnetic field, we come across magnetic field and magnetic field as continue to dive into cosmos. Hence it's quite natural that imprint of magnetic field must be manifest in functioning of our brain, our cognition and our language.
In recent works, [1], the present author studied natural languages and have found existence of a curve magnetisation under each language. We termed this phenomenon as graphical law. Then we looked into, [2], dictionaries of five disciplines of knowledge and found existence of a curve magnetisation under each discipline.

We have found, [1], three type of languages. For the first kind, the points associated with a language fall on one curve of magnetisation. For the second kind, the points associated with a language fall on one curve of magnetisation, once we remove the letter with maximum number of words or, letters with maximum and next-maximum number of words or, letters with maximum, next-maximum and nextnext-maximum number of words, from consideration. There are third kind of languages, for which the points associated with a language fall on one curve of magnetisation with fitting not that well or, with high dispersion.

Probable explanation for the second kind of languages, of the requirement of first maximum(say) of being removed, may be that many words associated with the very letter have not been assimilated (or, embedded) into the language in the form of complete cycles of activity. To illustrate, read, write, seminar, question, research, print, publish form one complete cycle of activity.

In English society, this activity is fully embedded. Coincidentally, we notice that words which form parts of this activity start with letters: p,q,r,s, w; are as close as possible.

Compare with related bengali words: pora,lekha, prosno, gobesona, prakasona. Two words are missing. ”La”, ”pa”,”ga” are far away in the bengali alphabet with the fact that though pora, prosno are paired quite well, lekha, prakasona, gobesona were not probably coupled among them or, with pora and prosno in the grassroot level.

When a language is strict in accomodating words which form sets of complete cycles of activity, we observe words following the points on a curve of magnetisation.

When a society cannot cope up with many complete cycles of activity, but has to go along with fractional cycle(s) of activity, the language of that society following the trend, imbibe words which are discrete, resulting in jump in number of words starting with a letter or, two, in dissonance with words beginning with other letters.
In this work, we continue our enquiry into the bengali language. Present form of bengali language got consolidated through the presence of Vidyapati and Krishnapada. The script currently in use had its final shape at the hand of Ishwar Chandra Vidyasagar.

We take a bengali to bengali dictionary, [3]. We try to see whether a graphical law is buried within the bengali language, in this article. The number of words counted by us for each letter is tabulated and graphically presented in the following.

I.1 Bengali

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2595	1397	177	35	1034	30	25	237	28	113	30	2314	599	1157	316	988	350	895	235	236

21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
137	191	134	1078	102	1392	515	1463	3196	392	3170	791	1773	356	737	434	955	47	2530	629

Here, the first row represents letters in the bengali alphabet in the serial order. The largest number of words, 3196 to be specific, start with the letter ”pa” probably pointing to its prakrit or, pali background. The next block of words numbering 3170 with the letter ”ba” as the initial, might be due to the fact that the name of the language is bangla. We draw number of words vs. sequence number of letters in the fig.1.

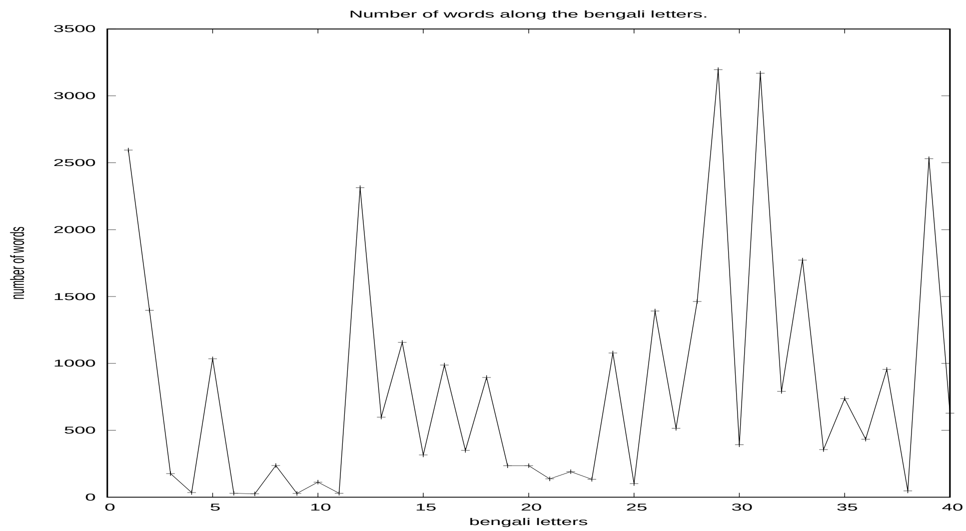

Figure 1: Vertical axis is number of words and horizontal axis is respective letters. Letters are represented by the number in the alphabet or, dictionary sequence. ([3]).

I.2 Magnetisation

Let us consider a coin. Let us toss it many times. Probability of getting head or, tale is half. If we attach value one to head, minus one to tale, the average value we obtain, after many tossing is zero. Instead let us consider a one-sided loaded coin, say on the head side. The probability of getting head is more than one half, getting tale is less than one-half. Average value, in this case, after many tossing we obtain is non-zero, the precise number depends on the loading. The loaded coin is like ferromagnet, the unloaded coin is like paramagnet, at zero external magnetic field. Average value we obtain is like magnetisation, loading is like coupling among the spins of the ferromagnetic units. Outcome of single coin toss is random, but average value we get after long sequence of tossing is fixed. This is long-range order. But if we take a small sequence of tossing, say, three consecutive tossing, the average value we obtain is not fixed, can be anything. There is no short-range order.

Let us consider a row of spins, one can imagine them as spears which can be vertically up or, down. Assume there is a long-range order with probability to get a spin up is two third. That would mean when we consider a long sequence of spins, two third of those are with spin up. Moreover, assign with each up spin a value one and a down spin a value minus one. Then total spin we obtain is one third. This value is referred to as the value of long-range order parameter. Now consider a short-range order existing which is identical with the long-range order. That would mean if we pick up any three consecutive spins, two will be up, one down. Bragg-Williams approximation means short-range order is identical with long-range order, applied to a lattice of spins, in general. Row of spins is a lattice of one dimension.

Now let us imagine an arbitrary lattice, with each up spin assigned a value one and a down spin a value minus one, with an unspecified long-range order parameter defined as above by $L = \frac{1}{N}\Sigma_i \sigma_i$, where σ_i is i-th spin, N being total number of spins. L can vary from minus one to one. $N = N_+ + N_-$, where N_+ is the number of up spins, N_- is the number of down spins. $L = \frac{1}{N}(N_+ - N_-)$. As a result, $N_+ = \frac{N}{2}(1+L)$ and $N_- = \frac{N}{2}(1-L)$. Magnetisation or, net magnetic moment , M is $\mu\Sigma_i\sigma_i$ or, $\mu(N_+ - N_-)$ or, $\mu N L$, $M_{max} = \mu N$. $\frac{M}{M_{max}} = L$. $\frac{M}{M_{max}}$ is referred to as reduced magnetisation. Moreover, the Hamiltonian due to Ising, [4], [5], [6], for the lattice of spins is $-J\Sigma_{n.n}\sigma_i\sigma_j - \mu B \Sigma_i \sigma_i$, where n.n refers to nearest neighbour pairs.

The difference $\triangle\epsilon$ of energy if we flip an up spin to down spin is $2J\gamma\bar{\sigma} + 2\mu B$, where γ is the number of nearest neighbours of a spin. According to Boltzmann principle, $\frac{N_-}{N_+}$ equals $exp(-\frac{\triangle\epsilon}{k_B T})$. In the Bragg-Williams approximation, $\bar{\sigma} = L$. Consequently,

$$ln\frac{1+L}{1-L} = 2\frac{\gamma JL + \mu B}{k_B T} = 2\frac{L + \frac{\mu B}{\gamma J}}{\frac{T}{\gamma J/k_B}} = 2\frac{L+c}{\frac{T}{T_c}} \tag{1}$$

where, $c = \frac{\mu B}{\gamma J}$, $T_c = \gamma J/k_B$. $\frac{T}{T_c}$ is referred to as reduced temperature.
Plot of L vs $\frac{T}{T_c}$ or, reduced magentiastion vs. reduced temperature is used as reference curve. In the presence of magnetic field, $c \neq 0$, the curve bulges outward. Bragg-Williams is a Mean Field approximation. This approximation holds when number of neighbours interacting with a site is very large, reducing the importance of local fluctuation or, local order, making the long-range order or, average degree of freedom as the only degree of freedom of the lattice. To have a feeling how this approximation leads to matching between experimental and Ising model prediction one can refer to FIG.12.12 of [6].
In an approximation scheme which is improvement over the Bragg-Williams, due to Bethe-Peierls, [5], reduced magnetisation varies with reduced temperature, for γ neighbours, in absence of external magnetic field, as

$$\frac{ln\frac{\gamma}{\gamma-2}}{ln\frac{factor-1}{factor^{\frac{\gamma-1}{\gamma}} - factor^{\frac{1}{\gamma}}}} = \frac{T}{T_c}; factor = \frac{\frac{M}{M_{max}}+1}{1-\frac{M}{M_{max}}}. \tag{2}$$

$ln\frac{\gamma}{\gamma-2}$ for four nearest neighbours i.e. for $\gamma = 4$ is 0.693. For a snapshot of different kind of magnetisation curves for magnetic materials the reader is urged to give a google search "reduced magnetisation vs reduced temperature curve". In the following, we describe datas generated from the equation(1) and the equation(2) and curves of magnetiastion plotted on the basis of those datas.

I.2.1 Reduced magnetisation vs reduced temperature datas

BW stands for reduced temperature in Bragg-Williams approximation, calculated from the equation(1). Bethe(4) represents reduced temperature in the

Bethe-Peierls approximation, for four nearest neighbours, computed from the equation(2). The data set is used to plot fig.2. Empty spaces in the table mean corresponding point pairs were not used for plotting a line.

BW	BW(c=0.01)	Bethe(4)	reduced magnetisation
0	0	0	1
0.435	0.439	0.563	0.978
0.439	0.443	0.568	0.977
0.491	0.495	0.624	0.961
0.501	0.507	0.630	0.957
0.514	0.519	0.648	0.952
0.559	0.566	0.654	0.931
0.566	0.573	0.7	0.927
0.584	0.590	0.7	0.917
0.601	0.607	0.722	0.907
0.607	0.613	0.729	0.903
0.653	0.661	0.770	0.869
0.659	0.668	0.773	0.865
0.669	0.676	0.784	0.856
0.679	0.688	0.792	0.847
0.701	0.710	0.807	0.828
0.723	0.731	0.828	0.805
0.732	0.743	0.832	0.796
0.756	0.766	0.845	0.772
0.779	0.788	0.864	0.740
0.838	0.853	0.911	0.651
0.850	0.861	0.911	0.628
0.870	0.885	0.923	0.592
0.883	0.895	0.928	0.564
0.899	0.918		0.527
0.904	0.926	0.941	0.513
0.946	0.968	0.965	0.400
0.967	0.998	0.965	0.300
0.987		1	0.200
0.997		1	0.100
1	1	1	0

We organise the paper as follows. We explain our method of study in the section II. In the ensuing section, section III, we narrate our graphical results. Then we conclude about the existence of the graphical law in the section IV. We continue with relevant discussion in the section V. The section VI is acknowledgement one. We end with bibliography.

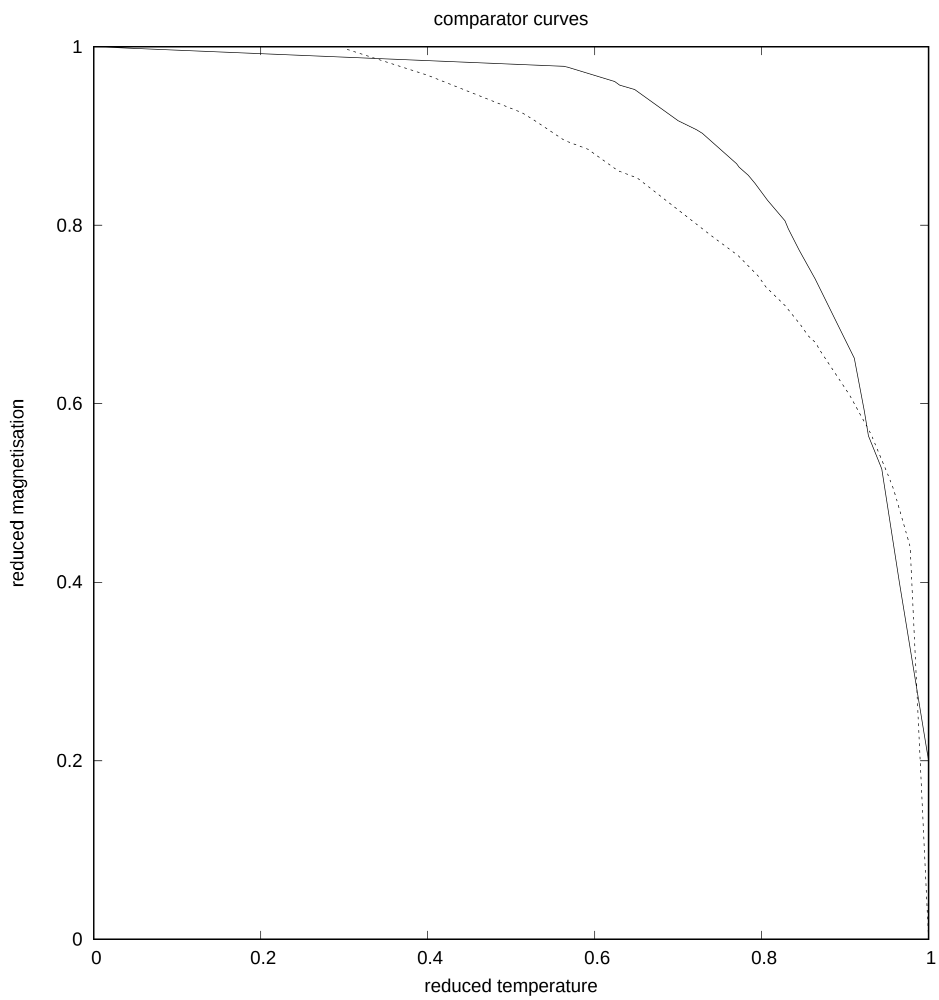

Figure 2: Reduced magnetisation vs reduced reduced temperature curves for Bragg-Williams approximation, in presence of little magnetic field and Bethe-Peierls approximation in absence of magnetic field, for four nearest neighbours (outer one).

II Method of study

We take the bengali-bengali dictionary,[3]. Then we count the words, one by one from the beginning to the end, starting with different letters. We assort the letters according to the number of words, denoted by f and the respective rank, denoted by k. k is a positive integer starting from one. Moreover, we attach a limiting rank, k_{lim}, and a limiting number of words. The limiting rank is maximum rank plus one, here it is forty and the limiting number of words is one. As a result both $\frac{lnf}{lnf_{max}}$ and $\frac{lnk}{lnk_{lim}}$ varies from zero to one. Then we plot $\frac{lnf}{lnf_{max}}$ against $\frac{lnk}{lnk_{lim}}$.

We then ignore the letters with the highest and then next highest and then next next highest number of words and redo the plot, normalising the lnfs with next-to-maximum $lnf_{nextmax}$, and starting from $k = 2$; next-to-next-to-maximum $lnf_{nextnextmax}$, and starting from $k = 3$; next-to-next-to-next-to-maximum $lnf_{nextnextnextmax}$, and starting from $k = 4$.

III Results

III.1 tables for graphs

k	lnk	lnk/lnk_{lim}	f	lnf	lnf/lnf_{max}	lnf/$lnf_{next-max}$	lnf/lnf_{nnmax}	lnf/lnf_{nnnmax}
1	0	0	3196	8.07	1	Blank	Blank	Blank
2	0.69	0.187	3170	8.06	0.999	1	Blank	Blank
3	1.10	0.298	2595	7.86	0.974	0.975	1	Blank
4	1.39	0.377	2530	7.84	0.971	0.973	0.997	1
5	1.61	0.436	2314	7.75	0.960	0.962	0.986.	0.989
6	1.79	0.485	1773	7.48	0.927	0.928	0.952	0.954
7	1.95	0.528	1463	7.29	0.903	0.904	0.927	0.930
8	2.08	0.564	1397	7.24	0.897	0.898	0.921	0.923
9	2.20	0.596	1392	7.24	0.897	0.898	0.921	0.923
10	2.30	0.623	1157	7.05	0.874	0.875	0.897	0.899
11	2.40	0.650	1078	6.98	0.865	0.866	0.888	0.890
12	2.48	0.672	1034	6.94	0.860	0.861	0.883	0.885
13	2.56	0.694	988	6.90	0.855	0.856	0.878	0.880
14	2.64	0.715	955	6.86	0.850	0.851	0.873	0.875
15	2.71	0.734	895	6.80	0.843	0.844	0.865	0.867
16	2.77	0.751	791	6.67	0.827	0.828	0.849	0.851
17	2.83	0.767	737	6.60	0.818	0.819	0.840	0.842
18	2.89	0.783	629	6.44	0.798	0.799	0.819	0.821
19	2.94	0.797	599	6.40	0.793	0.794	0.814	0.816
20	3.00	0.813	515	6.24	0.773	0.774	0.794	0.796
21	3.04	0.824	434	6.07	0.752	0.753	0.772	0.774
22	3.09	0.837	392	5.97	0.740	0.741	0.760	0.761
23	3.14	0.851	356	5.87	0.727	0.0.728	0.747	0.749
24	3.18	0.862	350	5.86	0.726	0.727	0.746	0.747
25	3.22	0.873	316	5.76	0.714	0.715	0.733	0.735
26	3.26	0.883	237	5.47	0.678	0.679	0.696	0.698
27	3.30	0.894	236	5.46	0.677	0.677	0.695	0.696
28	3.33	0.902	235	5.46	0.677	0.677	0.695	0.696
29	3.37	0.913	191	5.25	0.651	0.651	0.668	0.670
30	3.40	0.921	177	5.18	0.642	0.643	0.659	0.661
31	3.43	0.930	137	4.92	0.610	0.610	0.0.626	0.628
32	3.47	0.940	134	4.90	0.607	0.608	0.623	0.625
33	3.50	0.949	113	4.73	0.586	0.587	0.602	0.603
34	3.53	0.957	102	4.62	0.572	0.573	0.588	0.589
35	3.56	0.965	47	3.9	0.48	0.484	0.50	0.497
36	3.58	0.970	35	3.6	0.45	0.447	0.46	0.459
37	3.61	0.978	30	3.4	0.42	0.422	0.43	0.434
38	3.64	0.986	28	3.3	0.41	0.409	0.42	0.421
39	3.66	0.992	25	3.2	0.40	0.397	0.41	0.408
40	3.69	1	1	0	0	0	0	0

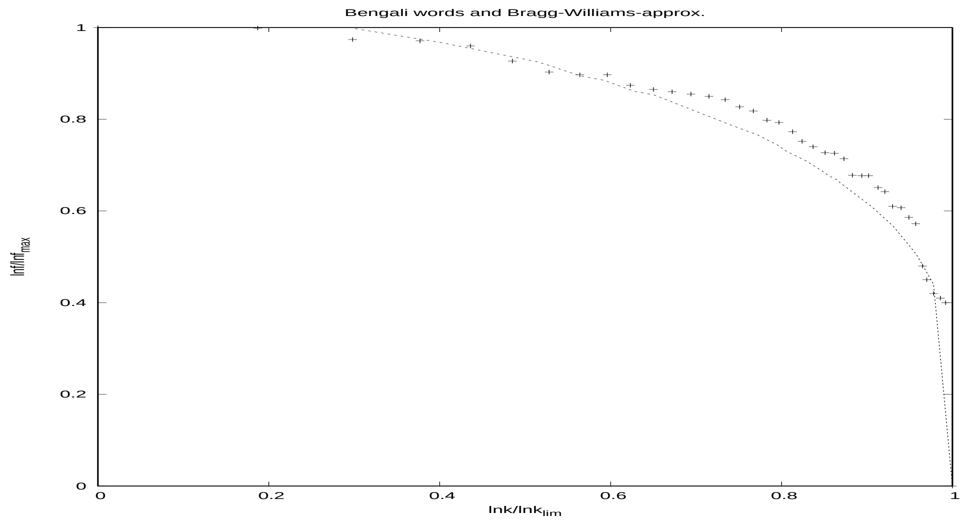

Figure 3: Vertical axis is $\frac{lnf}{lnf_{max}}$ and horizontal axis is $\frac{lnk}{lnk_{lim}}$. The + points represent the bengali words ([3]). Fit curve is Bragg-Williams line with little magnetic field.

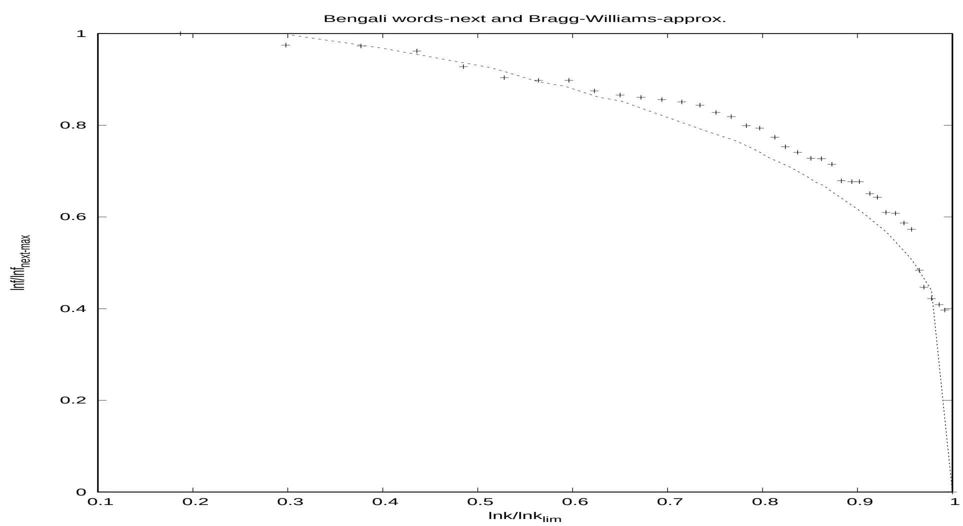

Figure 4: The + points represent the bengali words. Vertical axis is $\frac{lnf}{lnf_{nextmax}}$ and horizontal axis is $\frac{lnk}{lnk_{lim}}$. Visual match curve is Bragg-Williams line with little magnetic field.

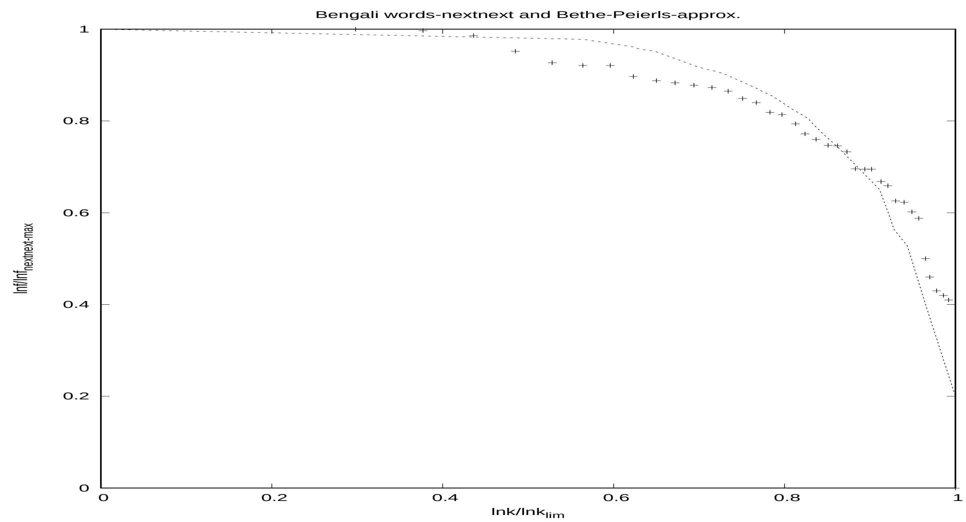

Figure 5: Vertical axis is $\frac{lnf}{lnf_{nextnextmax}}$ and horizontal axis is $\frac{lnk}{lnk_{lim}}$. The + points represent the bengali words ([3]). Comparator curve is Bethe-Peierls line for $\gamma = 4$ or, four nearest neighbours.

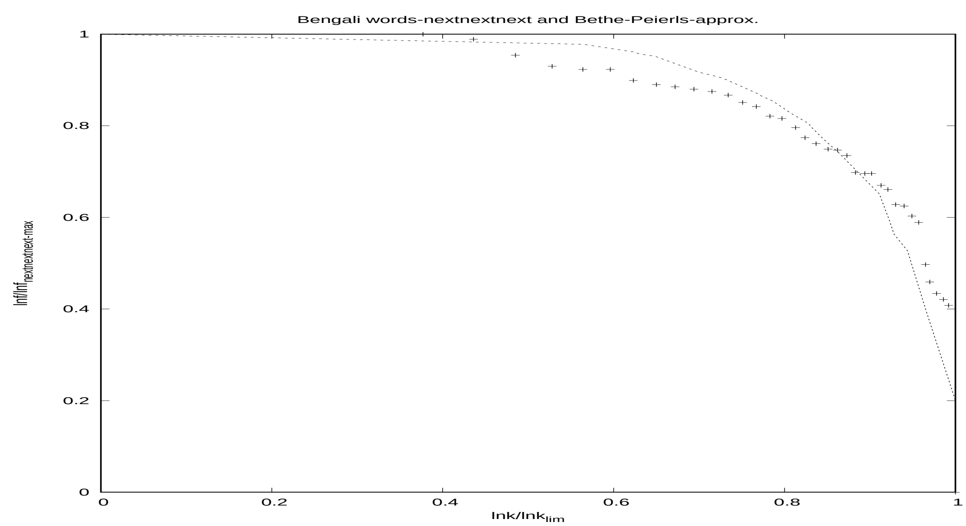

Figure 6: The + points represent the bengali words. Vertical axis is $\frac{lnf}{lnf_{nextnextnextmax}}$ and horizontal axis is $\frac{lnk}{lnk_{lim}}$. Best possible Comparator curve is Bethe-Peierls line for $\gamma = 4$ or, four nearest neighbours.

IV Conclusion

From the figures (fig.3-fig.6), we observe that dispersion is the least for the first figure i.e. $\frac{lnf}{lnf_{max}}$ vs $\frac{lnk}{lnk_{lim}}$ with the fit curve being Bragg-Williams line with little magnetic field.
The associated correspondance with the Ising model is,

$$\frac{lnf}{lnf_{max}} \longleftrightarrow \frac{M}{M_{max}},$$

$$lnk \longleftrightarrow T.$$

k corresponds to temperature in an exponential scale, [7]. Hence, bengali language follows graphical law. Bengali language can be identified with Bragg-Williams line with little magnetic field.

In bengal, "baro mase, tero parvan". Society was highly connected till a few decades back. As a result, society almost used to decide mode of life upto individual level. Reflection of this we see in the bengali language. Patterns of words are determined by society, innovation of words by individual(s) are rare to come by.

As temperature decreases, i.e. lnk decreases, f increases. The letterss which are recording higher entries compared to those which have lesser entries are at lower temperature. As bengali language expands, the letters like "pa", "ba", "." which get enriched more and more, fall at lower and lower temperatures. This is a manifestation of cooling effect as was first observed in [8] in another way.

V Discussion

To motivate the scenario, let us consider a collection of a set of people with similar interests. It may be a set of faculties or, a set of industrialists or, a set of students or, any other group. At zero temperature there is no random kinetic energy between two parts. These two parts may be a topshot and the rest in case of faculties; a successful innovative individual and the rest in case of industrialists. Then all the rest club together against(very rarely for) the singularity. Their jealousies align perfectly along the same direction. In the opposite situation, when everyone is performing at more or, less equal level, or, "no super and sub" then all jealousies get cancelled. A highly congenial climate prevails. Interaction is high. Exchange is high. "Temperature" is high. Hence, we can reasonably consider a lattice of faculties or, industrialists, or, families etc with an Ising model built on that. The correspondance is as follows:

$$jealousy \leftrightarrow spin,$$
$$coupling \quad between \quad different \quad units \leftrightarrow J_{ij},$$
$$level \quad of \quad collective \quad activity \leftrightarrow temperature,$$

Consequently, we get

$$\frac{collective \quad jealousy}{number \quad of \quad units} \leftrightarrow magnetisation,$$

and a curve of magnetisation is between $\frac{collective \quad jealousy}{number \quad of \quad units}$ and $level \quad of \quad collective \quad activity$.

Noun, verb, adverb, adjective and any word of a language is one or, other expression of jealousy. That's why we see underlying a language curves of magnetisation. People have time to be creative when activity is low. More number of words etc are generated then. If one set of people chooses one letter, another set of people chooses another letter, due to collective jealousy at higher scale of activity vs. collective jealousy at the lowest scale of activity. That's why see ranking of letters in logarithmic scale is anlogous to temperature and different arrangement of letters along the ranking for different languages.

VI Acknowledgement

We have used gnuplot for drawing the figures.

References

[1] Anindya Kumar Biswas, "Graphical Law beneath each written natural language", arXiv:1307.6235v3[physics.gen-ph]. A preliminary study of words of dictionaries of twenty six languages, more accurate study of words of dictionary of Chinese usage and all parts of speech of dictionary of Lakher(Mara) language and of verbs, adverbs and adjectives of dictionaries of six languages are included.

[2] Anindya Kumar Biswas, "A discipline of knowledge and the graphical law ", IJARPS Volume 1(4), p 21, 2014; viXra: 1908:0090[Linguistics].

[3] Rajsekhar Basu Sankalita(edited), "Chalantika, Adhunik Bangabhashar Abhidhan", thirteenth edition(teroso unonobboi bongabda (1982)), M. C. Sarkar and Sons private limited, 14, Bankim Chatterjee Street, Kolkata, © Dipankar Basu.

[4] E. Ising, Z.Physik 31,253(1925).

[5] W. L. Bragg and E. J. Williams, Proc. Soc. London Ser. A 145,699(1934); Kerson Huang, Statistical Mechanics, second edition, John Wiley and Sons(Asia) Pte Ltd..

[6] R. K. Pathria, Statistical Mechanics, p400-403, 1993 reprint, Pergamon Press,© 1972 R. K. Pathria.

[7] Sonntag, Borgnakke and Van Wylen, Fundamentals of Thermodynamics, p206-207, fifth edition, John Wiley and Sons Inc.

[8] Alexander M. Petersen, Joel N. Tenenbaum, Shlomo Havlin, H. Eugene Stanley, and Matjaž Perc, ”Languages cool as they expand: Allometric scaling and the decreasing need for new words”, Sci. Rep.2(2012) 943, arXiv:1212.2616v1. and references therein.