International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

The Economics of Digital Communication from Words to Bytes: Concern for Digital Leadership

Amrita Pratap[1], Vijit Chaturvedi[2], Ram Pratap[3]

[1]Ph.D. Scholar, Amiy Business School, Amity University, Noida, India
[2]Professor of OB and Psychology, Amiy Business School, Amity University, Noida, India
[3]Former Director, Indira Gandhi National Open University, New Delhi

Abstract

A simple yet robust model is presented for estimating redundancy in the English language, together with the significance of eliminating superfluous information, and repetition of messages / instructions. A survey of executives reveal that they do not like repetition of messages or notifications . Unlike Shannon's model for Information entropy and redundancy in English language, the proposed model accounting for 26 English alphabets is independent of text samples bias, and is not impacted with human variability. The redundancy in the language is found to be 62.3 percent which is calculated from the estimated value of information entropy 0.895 bit per letter. This figure indicates that a file can be compressed to about 38.7 % of its original size, aligning with the values of lossless file compression achieved with WAV and ZIP formats. Extrapolation of information entropy derived from our model to cover 256 characters of ASCII representation, yields a value of 8.77 bits per character. This value lends credence to ASCII encoding program which allocates 8 bits to each character, and to the LZW (Lempel -Ziv-Welch) dynamic encoding program which assigns about 9 bits to 12 bits respectively to frequently and infrequently occurring sequences of characters.

Keywords: Information Entropy, Redundancy in the Language, File compression, ASCII Representation

1. Introduction

1.1 Intelligent Digital Communication for Business Leaders

Digital communication has been gaining rapid momentum ever since it started, and is now virtually a lifeline of corporate communication which happens most often in the English language world over. Corporate leaders crucially depend on digital communication (online and offline) for sustained interaction with their teams, workforce, investors, and other stakeholders geographically scattered across the globe. A condensed and concrete message is not only a matter of intellectual liking but also a preferred format of an intelligent business communication. The impact and effectiveness of digital communication in any language can be significantly enhanced, and made economically efficient if digital leadership prioritizes, fostering proficiency in 'intelligent communication' . In essence, the intelligent communication is a holistic approach to conveying information that goes beyond mere transmission of information. It aims at a thoughtful and adaptable engagement with others, giving focus on the purpose of communication, cognitive level of receiver / audiences, tone and tenor in the text / speech, and the quality of the content in the message which should not be unnecessarily lengthy. Thus, the matter needs to be carefully curated and compacted, purging superfluous material.

1.2 Linguistic aspect in Digital communication:

"Speech is the representation of mind, and writing is the representation of speech." -
Aristotle

The communication (both speech and writing)- whether digital or in-person, is an interplay of wisdom, knowledge, information, and Language. Intelligent communication demands that the message is precise, concise, decisive, and free of redundant information. This not only makes the communication effective, but also helps in making the file compact. What is important in the message is the informational value that is to be retained and conveyed in the message, and perceived similarly by the receiver as conceived by the sender. Next, the overall size of the file can also be reduced further by using effective compression tools (utility software) available in the market, the selection of which will decide the quality of the compressed file reaching the receiver's end. While lossy compression tool can significantly reduce file sizes, but may cause permanent loss of fidelity or finesse, particularly in terms of visual or auditory quality in images-videos, or audio file; lossless file compression tools reduce file size up to a certain limit and the quality is retained. Conciseness and compactness of curated content , and overall compression of the file economize storage space, digital transmission time, and bandwidth in terms of employees productivity expense, hence the cost involved in digital communication.

Digital Leaders must, therefore, be cognizant of the quality of content , file compression software , file size, storage space, use of digital transmission time (DTT) , internet speed, and bandwidth limit that majorly affect the delivery of their messages. We briefly define the digital terms used:

File Compression involves the reduction of file sizes through encoding techniques, algorithms, and data compression methods. File compression is essential for optimizing storage space, reducing bandwidth usage, and improving transmission efficiency in digital systems. It contributes to the economic use of digital resources by minimizing storage costs and enhancing data transfer speeds.

Digital Time refers to the speed, efficiency, and management of time within digital systems and processes, particularly when big group virtual meetings are organised. This can include considerations such as data processing speed, network latency, response times in digital interactions, and the optimization of time-related resources in digital workflows.

Bandwidth economy involves optimizing data transmission, minimizing data congestion, prioritizing traffic, and maximizing the throughput of digital communications while minimizing costs. It plays a crucial role in ensuring the smooth functioning of digital networks and services while minimizing operational expenses.

1.3 The economics of digital communication:

The cost-effectiveness and value addition are the cornerstones of all business endeavours where 'optimum investment of time' is also a hidden but a crucial component of business to impact both cost and value addition. Since time is at premium in today's world, productive minds prefer to receive and send concise and concrete information with no extraneous material, which saves their engagement time. Compact and concise content, and compressed file serve both the purposes. While making information concise depends on the skill and ability of the communicator, the overall file compression (lossy or lossless) depends on file compression tool. The thought-provoking idea used in file compression encoding tool (software) is that it leverages the intrinsic property of the language – the 'information Entropy' or 'Shannon Entropy' The value of information entropy enables one to calculate the redundancy in the language, another intrinsic property of the language, which in turn depends on the linguistic factors, such as average word size, syntax rules, grammatical constraints, and the construction of the sentences. Compact and concise matter, and compressed files require less storage capacity, less time for file transfer while consuming less bandwidth, resulting in cost saving.

Shannon Information Entropy: In the context of language, Shannon information entropy quantifies the average information contained in a letter or character in terms of bits, or in other words, it shows bits required for a letter or a character to digitally communicate. Shannon's approach considers the probabilities of occurrence of different letters/ symbols in a general text. This is influenced by factors the such as word frequency, syntactic structures, and vocabulary richness.

Redundancy in language: In the context of information theory and language, redundancy refers to the degree to which elements of a message are predictable or can be inferred from looking at other elements. It is essentially the opposite of entropy. Redundancy arises from patterns, structures, and regularities inherent in syntax, grammar, and semantics, and allows for error detection, helping in effective communication.

Similarly, the average word size can impact both entropy and redundancy, as longer words may carry more information but also potentially introduce more redundancy through contextual predictability. Therefore, for understanding these subtle properties of a language, one needs to probe the linguistics aspects of the language.

2. New model- An overview:

Software tools that effectuate data compression utilise redundancy in the language , an inherent property of the language in use. When a system functions with a set of rules and constraints, it has a fixed format, hence its behaviour is more predictable. Curiously so, the systemic format, syntactic structure and grammatical rules of the English language build redundancy in itself that lends support to predict or guess the missing or illegible part of the printed text, and the interrupted segments of verbal communication even in a noisy background. Thus, redundancy in the language is leveraged for the compression of a data file, which can be done by a trained software. How much information in terms of bits is contained in a letter of English alphabet, and how large the redundancy in the English language is, was studied and explained by Claude E. Shannon [1] who first coined and defined the terms 'information entropy' and 'language related redundancy'. He formulated the equations to determine their values as a function of frequency counts of letters/ words. Redundancy in a language refers to the inherent property of a language, determined by the information entropy and its maximum value.

Thermodynamic entropy and Information Entropy

To gain quick understanding of how classical concept of thermodynamic entropy was extended to the field of digital communication, one needs to know in brief how entropy, probability, and information are interrelated. We walk through these terms and get to understand them qualitatively without going into the mathematical details here. If the entropy of a system of a particular configuration, and the probability of the system to assume that configuration are related, and if the probability of the configuration as referred to and the informational value attached to the configuration are related , then by simple logic, entropy and information must also be interconnected [2,3].

The term "entropy" was originally introduced by Rudolph Clausius in his work on the Second Law of Thermodynamics [4,5]. Thermodynamic entropy is a measure of the amount of thermal energy in a thermodynamic system that will not be available for doing useful work, and goes as a waste. The change in entropy of a system can be calculated as the change in the heat content of the system divided by its absolute temperature at which the work is being done on or by the system.

In the years to follow, Ludwig Boltzmann [6] gave the following entropy equation:
S = K(B) log W (E) ....(1),
where S is the entropy, K (B) is the Boltzmann constant, and W is the number of all possible configurations (microstates) or arrangements across which all the constituents (molecules) could be distributed in different ways under the condition, where all the configurations are equally probable and are consistent with the observable macro-state of the system having total energy content E.
J.W. Gibbs later generalised Boltzmann's idea, and characterised different configurations (micro-states) with varying probabilities [7,8]. The reformulated Boltzmann equation is:
S = - K (B) Sum (i = 1 to n) (pi log pi), ...(2),
where the summation is done over all the configurations / microstates, pi is the probability of configuration 'i', and K (B) is the Boltzmann constant. Janes [7] says, "Disorder is a manifestation of the largeness of the number of microstates the system can have. The larger the choice of microstates, the lesser the predictability or level of order in a system." Or to put it in other way, the lesser the certainty in defining the parameters of a system, more is the disorderliness present in the system.
The idea of entropy was further expanded to digital communication. Claude Shannon's groundbreaking work on information theory, published in 1948 ("A Mathematical Theory of Communication") [9] and expanded in 1951 ("Prediction and Entropy of Printed English") [10], laid the foundation for measuring information entropy and quantifying the amount of information contained in a message in terms of the binary digits (bits), which laid the foundation of Information digital Communication. He estimated the values of information entropy and redundancy in the English language.
When the language is translated into binary digits for storing, and transmitting over a channel, the information entropy is expressed in bits per letter / character which implies how many bits are required to store a letter or how much information in terms of bits is generated by a letter or any other character. The transmission of a message over the channel is given in bits per second, called bit rate. Shannon presented Information Entropy H (X) of a discrete variable X (letter or character) as a function of probability given in the following formula:
H (X) = - Sum (i = 1 to n) pi log (base 2) of (pi) .... (3),
where the sum is taken over all letters (i = 1 to n); the logarithmic base is 2 and the unit of information entropy is bits per letter or character, and pi is the probability of a letter or character 'i' appearing in the string of letters/characters. The information entropy of an English letter refers to an average amount of information (minimum number of bits) that one can obtain from the letter in the context of a language model, or the minimum number of bits required to represent a letter/ variable.
Shannon's formula for redundancy (R) is:
R = 1 - H (X) / H (X)(max) ....(4),
where H (X) (max) is the maximum value of the entropy associated with a variable/ letter when all variables / letters are equiprobable, or have uniform probability distribution in the probability space.
In terms of the outcomes of an event, the upper bound of entropy is reached when all possible outcomes are equally likely, whereas the lower bound of the entropy is reached when one outcome is certain to occur and all other outcomes have very low or near zero probability. This implies that if all the variables are distributed uniformly across the sample, the entropy reaches its maximum value, and in this condition, it is the same as the upper bound of entropy.

One way to assess information entropy qualitatively is by looking at the frequency of letters or words in a text. A text message with high degree of randomness / irregularity in its word- distribution will have higher entropy, while a text with low degree of randomness or less irregularity of its word-distribution will have lower entropy. Kwiatkowski [11] says, 'Entropy is a measure of uncertainty or randomness in a system. High entropy means high uncertainty.' Here, he connects entropy and irregularity with the uncertainty. Let us understand it in a simple way: if a system is irregularly or randomly behaving or responding or functioning, its future response or behaviour will also be uncertain, unpredictable, or unguessable. And if any system is behaving periodically with a certain pattern over time , it future behaviour will also be predictable / guessable. In case, some event / disclosure happens which was not predictable or guessable , it is a surprising outcome that is to be associated with high entropy. A few simple examples would clarify it. The cities of Birmingham (United Kingdom) and Lubbock (Texas, U.S.) have most uncertain weather which changes suddenly anytime; but if weather stays calm for long time, it is surprising. The collapse of Lehman Brothers' Holdings Inc- onetime the largest investment bank- in 2008 was not at all expected, still it happened and triggered global financial crisis ; it was surprising. This indicates that such surprising outcomes are associated with high entropy.

Cover and Thomas [12] provide a comprehensive introduction to the fundamental concept of information entropy and data compression. They discuss how entropy can be used to characterize the statistical properties of a random process, as well as how it can be used to measure the amount of information contained in a letter or string of letters (message). Whereas the information entropy in bits per letter measures information contained in or produced by a letter or a word, language's intrinsic redundancy measures the extent to which the text of a language could be predictable or guessable. Language's intrinsic redundancy resides in the very structure of the language and is mathematically derivable from the information entropy of the language. Since the syntax of English is governed by certain rules of grammar, tenses, and parts of speech, the structure of the language itself works as an enabler to comprehend the continuity of a text, or a sequence of letters to complete a meaningful word or full text. Accordingly, it provides cues to fix unfilled /undecided spellings of letters or incomplete sentences by way of guessing what needs to be added or placed next to the truncated string of given letters/words /sentences. Thus, a person even with an average knowledge of the English language can guess deleted or missing letters in a word, or missing words in phrases or sentences, or scribbled parts of sentences in a text, and can even guess interrupted chunks of a speech (auditory communication) amid the babble of conversation. This makes the basis of auto spelling correction software and Artificial Intelligence.

Shannon's theoretical and experimental approach:

Shannon used N-gram language model for the estimation of information entropy of the English language, which estimates the frequency of occurrence of Nth letter in relation to the occurrences of previous N-1 letters. As N increases to letter 3 -gram, the value of information entropy approximates to 3.3 bits per letter. Shannon also calculated the frequency counts of each word of a corpus of over 8000 words, and obtained the weighted average of information entropy 11.82 bits per word. With the average word length of 4.5 letters he used, the entropy per letter turns out 2.62 bits, and the corresponding redundancy in the language 44.3 percent. From an experimental study on human subject, with average good knowledge of English, based on guessing letters/words in the sentences of different paragraphs of English texts, Shannon found that the subject made 69 percent correct guesses, and 31 percent incorrect guesses. This would mean that by having to know only 31 percent of the information, the human subject could still get right understanding of the whole lot of the original text, indicating, therefore, 69 percent language redundancy;

the reverse calculation using Shannon's redundancy formula yields information entropy 1.45 bits per letter. What is observable from the results of different procedures Shannon adopted is that the information entropy value varied from 1.45 bits to 3.3 bits per letter. Further, when the string of characters increased from 8 to over 100, the statistical fluctuation ceased and entropy reduced to the order of 1 bit per letter, and redundancy rose to 78.7 percent. While Shannon's theoretical approach required approximation and involved uncertainty beyond letter 3-gram, his guess - experiment on the human subject too suffers on two counts: the first is that all the passages of English text were selected from the same book, and the second is that all responses came up from only one respondent.

Burton and Licklider [13] examined the statistical structure of printed English with a focus on long-range constraints: how long-range string of letters influence the organization of words and sentences, and how they can be used to predict the probability of certain words or phrases occurring in each text. Dolby and Resnikoff [14] also discussed various factors such as etymology, pronunciation, and grammatical framework that affect the structure of the English language.

Theoretical Framework of New Model, and Data Compression:

Data files can be compressed using a compression algorithm. The reduced size of the compressed data file requires less computer - storage and less bandwidth to transmit the data. As a result, it saves digital transmission time to certain extent, and makes storing or transporting of data more efficient, and cost effective. Compression algorithms can work on any type of data, not just words. Compression algorithms are designed to take a piece of data - a text, image, audio, or video- and encode it in a way that uses fewer bits than the original data. A compression tool identifies repeating patterns in the data and encodes them using a more efficient representation. When the data is decompressed at the receiver's end, those patterns are decoded and placed back into their original positions in regularity, so that the resulting data is as much similar as possible to the original input data. The decompression process essentially reverses the compression process, using the encoded representation of the data to reconstruct the original data as closely as possible. In most cases, the decompressed data/ file is expected to be identical to the original input data, though there may be some differences due to the compression process executed by a coding algorithm of a particular software format. Output should be ideally loss free; and if it is lossy, then it should be at minimal loss of information that meets the requirement of both the sender and the receiver. In auditory communication, the voiced region has low entropy due to the set format of the language and set formants in the acoustic spectrum which, by and large, are seemingly not affected by the punctuation marks. Phonetic transcription would not always match with phonemic transcription which, of course, matters a lot in speech recognition. Pierreumbert [15] argues that phonology and phonetics are two different levels of representation in the grammar of a language. Phonetics is concerned with the physical properties of speech sounds (frequency, amplitude, intonation, and articulation) whereas Phonology is concerned with the way phonemic sounds are organized and integrated with a language to convey meaning. It is important to understand that, in the English language, one letter can represent more than one sound or may even remain unsounded as in French. All such studies helped make rapid advances in the fields of AI and NLP, and opened a vast scope for research in these fields.

The file compression is related to in-built redundancy in a language, which is derivable from the 'information entropy' that itself arises owing to the constraints of grammar, syntactical rules, structural requirements for a sentence of a particular language.

Procedure adopted in the New Model:

Our study presents a novel model and method for obtaining accurate, verifiable, and consistent values of the information entropy, and entropy- based redundancy in the English language. In the organization of a message, it is also important to eliminate or at least minimize man-made superficial information which accumulates due to the repetition of words, phrases, and sentences because it adds unnecessary digital load to files and may impede listener's cognitive process. We discuss it later in another section. The model encompasses all the 26 letters of the English alphabet. It does not differentiate between upper and lower case, and does not include characters such as punctuation marks, or other symbols that are inactive or have limited value in auditory communication (voice or videos). However, these characters have a guessable pattern of repetition in a written text, hence it easier to encode and decode them.

The model focuses on the composition of an average English word, and examines the relative frequency counts (probability of occurrence) of vowels and consonants present in an average word-length to estimate the value of information entropy. Knowing the relative frequency counts, we apply Shannon's formula (refer equation 3) to estimate information entropy. This approach is different from Shannon's original approach.

Another departure in our model from Shannon's method comes in the estimation of maximum value of information entropy which Shannon used to calculate redundancy (equation 4). The present analysis deals it with a challenge since the letters in an English word belong to two distinct subgroups (vowels and consonants) of English alphabet, having different population, hence cannot be treated alike for equal probability distribution required to estimate maximum value of information entropy. We, therefore, treat the letter space and probability space of each subgroup separately, and then apply equiprobable condition independently. However, in the construction of a word, both the subgroups have equal probability to introduce their letters to build a word. Thus, considering first the equal probability of the subgroups to introduce a letter in a word, followed by the equal probability of the letters from the respective subgroups to join a word, we calculate the compound probabilities and substitute them in Shannon's entropy formula to obtain the maximum value of the information entropy for vowels and consonants separately. Subsequently, by assigning weight factors to these values based on their frequency counts in an average length of an English word, we take the weighted average which serves as a true representative of the maximum value of entropy per letter. The results of our model are compared with Shannon's refined or average values.

In the calculation of probabilities and estimation of information entropy, one needs to know the average length of an English word and the number of vowels and consonants it contains. This inquiry provides an intriguing insight into the evolution of the English language, which has been shaped by the influences of various other languages, including Greek, Latin, French, German, and Spanish. During its early development, English borrowed heavily from these languages, incorporating words with or without modification. Over time, the intervention of linguists and etymologists became more prominent as they applied orthography-based analytics [16] in the construction of words. Spoken words were transcribed into written text based on their pronunciation, often with inconsistent relationships between sounds and spellings. Miller [16] explores this relationship between written and spoken English, highlighting that a single word may have similar letters with different sounds, or different letters with similar sounds, or may even remain unsounded.

The composition of words in fiction, stories, novels, and literary or scientific works in different eras of history had been changing the average length of an English word. Bochkarev et al. [17] conducted an analysis of the dynamics of the average word-length in English and Russian, and the impact of society on the use of function words and content words. The study found that from 1901 to 1994, the average word

length of English increased linearly to reach 5.2 letters, and then started to slightly decrease from 1995 to 2008. It was also noted that the average word length in American English was different from British English. However, more comprehensive data seem to have emerged later.
Mark Davies [18], a Professor of Linguistics at Brigham Young University, compiled a corpus of 5,000 commonly used words and studied several cases of syntactic variation in English. Leo Qin [19], Manager at Texas Health Action, conducted a computational analysis of the same corpus and found that the average length of an English word has risen to 6.455 letters, with an average of 2.029 vowels and 4.426 consonants, while the mean value of average length is 6. The average value of 6.455 provides a more accurate representation of the typical word- length since it accounts for reasonable variations but ignores abrupt drop or spike; the mean value, however, is influenced by outliers.

Results and Discussion

Having followed the procedure to obtain the probabilities of occurrence of vowels and consonants in an average English word, Shannon's formulae are applied to calculate (Appendix) the values of information entropy, and redundancy in the English language, which could also be termed as entropic redundancy because it depends on the information entropy. The model yields the value of information entropy 0.895 bit per letter, and entropic redundancy of the 62.3 percent, based on an average word length of 6.455 letters. When the mean value of 6 letters word- length with 2 vowels and 4 consonants is used in the calculation, the information entropy and language redundancy change slightly to 0.916 bit per letter and 61.1 percent respectively. As the model does not depend on human variability which accounts for the cognition and speech characteristics, the model provides the least -biased estimate of information entropy and language- entropic redundancy both for written and spoken English. In the printed text, some punctuation marks are frequently used but they all follow repeatedly a set pattern, hence can be easily encrypted with a trained software, and then can be decrypted at the receiver’s end. If need arises, the model can be modified to incorporate other characters and numerals according to their relative importance.
In the following subsections, we compare our results with the findings of Shannon, and other authors, as well as the compression algorithms in use. The analysis with our model yields information entropy 0.895 to 0.916 bit per letter and the corresponding redundancy 62.3 to 61.1 percent. Thus, the information entropy obtained from our model approximates to 1 bit per letter and is understood to be free from statistical- fluctuation because it matches closely with Shannon’s result for the long chain of characters. Shannon's claim that the statistical fluctuation ceases for a long sequence of over 100 characters, has been disputed by other authors as they got this happen around 31 characters, beyond which the value of entropy remains unaffected [21].

Comparison with LZW algorithm, ASCII representation, WAV and ZIP formats:

The Lempel-Ziv (LZ) algorithm [22] is a family of data compression algorithms developed by Abraham Lempel and Jacob Ziv in 1978. The Lempel-Ziv algorithm works by identifying repeated patterns in the data and encoding them using shorter codes. It builds a dictionary of patterns encountered in the input data and uses references to previously seen patterns to represent subsequent occurrences. By replacing repeated patterns with shorter codes, the algorithm reduces the overall size of the compressed data. Terry Welch [23] improved Lempel-Ziv algorithm by using a specific technique of creating dynamic dictionary which keeps checking and learning in real-time from the streaming input data. Lempel -Ziv-Welch (LZW) algorithm [23] works with average code length varying from 9 to 12 bits per character: longer sequence and/or frequently occurring sequence of characters are given shorter codes, whereas the shorter sequence and/or less frequently occurring sequence of characters are given longer code. By assigning shorter codes to more frequently occurring character sequences, and longer codes to less frequently occurring character sequences, the algorithm can achieve greater compression ratios, consuming less digital storage. The algorithm assigns variable-length codes to input data for reducing its entropy or, to say, the data file size. In a sense, the varying code length in the LZW algorithm can be treated as a measure of the information entropy of a language. The result is that the compressed data uses overall fewer bits to represent the original data, making it more efficient to store or transmit.

If we extrapolate our model to include 256 characters of extended ASCII's coding list and then work out information entropy, we find the value reaching around 9 bits per character. This is strikingly an important result because ASCII representation assigns 8 bits to each character whereas the average code in LZW algorithm [23] works around 10 bits per character.

Thus, our model which is independent of approximations and human subject variability, is supported by ASCII 8-bit representation, as well as by average 10-bit code length of LZW compression algorithm. Using structural nuances and rules of the language for encoding with special software tools, standard ZIP format can compress word files by 62 % and retains the qualities of the original file. MP3 can shrink the digital audio file to approximately 85% but it is a lossy compression. WAV format can compress an audio file by 60 % of its original size, and this compression is understood to be theoretically lossless. Our model, which is independent of text sample or audio sample, gives language- entropic redundancy 62.3 and 61.1 percent. This would mean that word doc or audio file can be compressed by 62.3 or 61.1 percent which are close to the values ZIP and WAV formats offer for lossless compression.

Redundant Information and survey on Executives:

“The single biggest problem in communication is the illusion that it has taken place.”

-George Bernard Shaw

Redundant information (superfluous information) in digital communication According to the Oxford Dictionary and Thesaurus, the word redundant is defined as something superfluous, not necessary or something unnecessarily stated again and again.

Redundant information in digital communication does not depend on the information entropy, and has nothing to do with the information entropy-based redundancy in a language. Redundant information is the superfluous usage of words or phrases having similar meanings which can be spotted in a shoddy communication and the way language is casually and inefficiently used to convey information. It can be seen in the passages of a poorly drafted word document, e-mail, or unprofessionally organized audio or video files. The communicator is primarily responsible for redundant information though this too can be controlled, to a certain extent, by compression logarithm. Still, lesser the redundant information, better it is. Redundant information does cause wastage of digital- space and digital- time, and may cause distraction in reader’s /listener’s cognitive process. As a result, the intended impact of the message and purpose of communication is diminished. Hence, what is significant here is the knowledge of semantics and the skill and ability to properly organise information before it is presented in a discourse or transmitted over the channel.

If redundant words/sentences are reduced to minimum or eliminated, the data file itself becomes concise and saves digital space. This would help in the effective utilization of channel bandwidth and enhance the quality of communication. Unneeded repetition of information in the same text or speech seems crude though often mistaken by the sender as persuasive for making a deep impact of communication. Such stuff is a kind of noise or an unwanted bunch of random frequencies that distort the signal and corrupt the files also. This can also adversely affect the interpretation of the message and can lead to infructuous altercation with the audience. Mandelbrot [24] discusses how the meaning of words is often determined by the context in which they are used. He also unfolds how the structure of a message with proper syntax enhances the efficiency of communication and optimizes the transmission time of information. Those who are habitual of using redundant text or information are required to emend their documents or speech wisely. Sperber and Wilson [25] have noted that listeners are cognitively demanding and barter minimum effort for maximum cognitive experience - intellectual, expositional, and motivational. This is a contradictory min-max expectation, still it has reason to be met by the leader. According to Relevance theory [26], communication is relevant to an individual if one can find positive cognitive effects from the processing of the received message. Relevance means whether communication shows truthfulness and has clarity of thoughts; whether the matter is contextual and timely important; whether the utterances are based on facts, and carry informational value. Lack of awareness of these facts can lead to the failure of business ventures and breakdown of inter-personal relationship. Digital leaders need to be aware of this fact.

The development of digital culture [27] in the organisation depends on digital leadership. Therefore, leadership must be versatile [28] to simultaneously meet the requirements of the company, and the demands of markets. The specifics of digital leadership/ e-leadership have also been given in detail by Torre and Sarti [29]. They discuss how patterns of leadership are changing with the support of technology. Leadership must be conscious of emerging technologies and supportive of the need for new norms to manage workers and executives effectively. Unless conscious efforts are made, leadership might not notice that there is a hidden flip side of technology too. The emerging technologies are creating a venetian blind between the team leaders and their team members; leaders are gradually losing human touch with their team, and creating a virtual space for confusion and criticism. Effective digital communication, therefore, poses a challenge to digital leaders who are supposed to encourage digital culture, manage the company's digital assets, and organize efficient and effective use of language for digital mediums and platforms. Leaders in general must empower their team not only with the changing landscape of technology that impacts the business world, but also with face-to-face interaction and their physical presence, whenever and wherever possible.

Managers averse to repetition of information or/ circulars / notifications:

In a recent study conducted on different target groups, including primary and middle-school going children, skilled and semi-skilled workers, and executives and managers of medium to large size organizations, the effectiveness of repetition of instructions or notices or circulars was evaluated. To elicit responses from children, we contacted many of their parents also. The findings of survey reveal that the preference for repetition of instructions varies across different target groups. Most children prefer more repetitions of their lessons and reminders for the completion of their tasks. The responses from the semi-skilled workers are also more or less the same. They all feel that it puts them on alert and motivates them to perform. In contrast, the opinion on the repetition of instructions among skilled workers was divided almost fifty-fifty. However, in the case of executives and managerial class, more than 70% of managers we contacted, stated that the frequent reminders do not enhance their efficiency or effectiveness. " Repeated messages, instructions, circulars, and notifications issued by leaders or seniors now and then on the same matter become boring, irritating, and unappealing to the recipients ", stated corporate managers in our face-to-face informal discussion.

Professor Neeley of Harvard Business School [30] recommends repetition of instructions for children. The basis of the argument is that the repeated instructions given to children by their parents deliver positive results. Children then complete the tasks they are asked to perform, such as brushing their teeth, going for a shower, and taking breakfast, etc. She further suggests that repeated reminders can also work for the corporate executives to get their projects completed quickly. It is important to note that applying child psychology to adult executives working in corporate appears to be inappropriate, as the behavioural psychology of children under the parental influence and guidance, and those of adult executives with experience, interacting with peer group, and having independent aspirations in life are poles apart. Moreover, the reasons for delay in project completion or implementation are most often due to the hurdles in coordination between different divisions, lack of timely input, and lack of infrastructural support, rather than any lacking on the part of the manager handling the project. Therefore, the impression that frequent reminders can make executives work efficiently, and the onus for the delay can be transferred to them is an evasive and mistaken view of seniors. It is important for leaders to understand the underlying reasons for the delays and take appropriate measures rather than relying solely on reminders alone.

Conclusion:

The proposed model is simple and versatile; it considers all 26 letters of the English alphabet, excluding all other characters. The letters found in an average English word are classified into two subgroups: vowels and consonants.

The information entropy per letter is estimated by separately examining the frequency counts of vowels and consonants in the average length of an English word. To estimate the maximum value of information entropy as required in Shannon's formula for redundancy in the language, the condition of uniform probability distribution is applied separately to both the subgroups of vowels and consonants in their respective probability spaces. Then, the weighted average is calculated by assigning the weight factors individually to the values calculated separately.

Further, we obtain redundancy value of 62.3%, which matches with value reported by Moradi et al. [21] who worked with different text samples and human subjects. These findings corroborate that our model is not affected by the variables like text samples or human subject. In Shannon's experiment, when the string of characters is increased from 8 to over 100, the short-range statistical fluctuations cease, and the entropy drops to around 1.0 bit per character, which is very close to the value of 0.895 bit per letter obtained from our model. This suggests that the present model is resistant to statistical fluctuations.

Unlike other models and approaches, our model does not have any theoretical or experimental constraints. The results obtained from our model are not based on any approximation, are unaffected by the length of the string of characters, and are free from self-selection bias of test samples and human variability. Therefore, the model is more robust, reliable, and objective

File compression of 62.3 per cent, as inferred from our model, is compatible with the lossless compression provided by the WAV and ZIP file compression formats. When the entropy value obtained from our model is extrapolated to accommodate 256 characters, the resulting value goes to 8.77 bits per character. One may recall that ASCII character encoding works with 8 bits per character, while the LZW algorithm uses code lengths around 10 bits for individual character or groups of characters. Thus our results are corroborated by ASCII representation and LZW algorithm.

Redundant information in speech or written text often arises from the repetition of words and phrases with similar meanings or the repetition of sentences that are illusively perceived as persuasive. This essentially happens due to the lack of preparation, limited knowledge of semantics, and uncontrolled emotional outbursts. Redundant information not only wastes digital space and time in terms of employee's efficiency expense, but also disrupts the cognitive processes, eventually diminishing the intended impact of communication.

While language plays a pivotal role in all dimensions of communication (social, business, and strategic), digital communication - written and spoken – has become an indispensable mode of knowledge-sharing and information- exchange across the globe. The rapid advances in Internet technology have led to the democratization of knowledge and instant dissemination of information, thus empowering the people of all strata of the society. Google states their mission:

'To organize the world's information and make it universally accessible and useful.'

It is good. But this freebie offer can turn into a destructive weapon if handled by the wrong mind. Hence, the acquisition of knowledge to understand the nuances of language while organizing and transmitting the information, and the application of wisdom to assess the implication of message invite special attention and awareness on the part of group leaders.

Appendix

Calculation of Information Entropy, its Maximum value, and Language Redundancy

Leo Qin's data has been used in the framework of our model. Out of 6 .455 letters word- length, 2.029 are occupied by vowels, and 4.426 by consonants. Hence, the probabilities of vowels and consonants located in an English word are:

p (vowels) = 2.029 / 6 .455; p (consonants) = 4.426 / 6.455. By substituting these values in Shannon's formula for Information entropy (Shannon entropy), we get:

H (X) =
{(2.029 / 6.455) log (base 2) of (6 .455 / 2.029)} +
{(4.426 / 6.455) log (base 2) of (6.455 / 4.426)}. .... Eq. (1)
= 0.895 bit per letter ……… 1(a)

If we calculate the entropy using a mean value of the word length as 6, and using the rounded-off number 2 for vowels, and 4 for consonants, the information entropy obtained from the same formula is
= 0.916 bit per letter. ……….1(b)

Now to calculate the language redundancy R, we need to know H (X) (maximum), the maximum entropy of a discrete variable X. It should be remembered here that sample space and the probability space of both the subgroups of variables- vowels and consonants – are different. There are 26 letters in the English alphabet where the sample space of vowels consists of 5 vowels (a, e, i, o, u), and the sample space of consonants consists of the remaining 21 consonants. The maximum information entropy H(X) (max) for 5 vowels, and for 21 consonants would depend on the equal probability distribution of the letters separately in their respective probability spaces. Thus, equal probability distribution for letters in the subgroup of vowels would be
(1/5,1/5,1/5,1/5,1/5), and for the letters in the subgroup of consonants, it would be 1/21, 1/21, and so on till we cover all the 21 consonants.

Initially, both the subgroups have equal probability, 1/2 and 1/2, to dispense the letters for the construction of an average word. Thereafter, the vowels and consonants from their respective probability spaces have again equal probability of occurrence in the word. This would mean that each vowel has a probability of 1/5 and each consonant has a probability 1/ 21. Thus, H(X) (max) for vowels is calculated with compound probability 1/2 * 1/5 = 1/10, while H(X) (max) for consonants is calculated with the compound probability 1/2*1/ 21 = 1/42. Hence,

H(X) (max) (vowels) = {5* (1/10) log (base 2) of 10} ....Eq. (2)
= (1/2) (3.321) = 1.660
H (X) (max) (consonants) = {21* (1/42) log (base 2) of 42} ....Eq. (3)
= (1/2) (5.392) = 2.696

It is to be noted further that the occurrence of all vowels, and all consonants is independent of each other, hence the uncertainties associated with the letters of both the subgroups are additive, and so are the entropies. We should also recall that the letters – vowels and consonants- building up an average word have only limited and definite number of sites to occupy. Therefore, we calculate the weighted average of the both the values of maximum entropy, equations (2) and (3), by assigning them weight factors in terms of their respective frequency counts, as given below:

H(X) (max) (weighted) = [1.660 * 2.209 + 2.696 * 4.426] / 6.455 … Eq. (4)
= [3.368 + 11.932] / 6.455 = 2.370

Substituting the values H(X) = 0.895; and H(X) (max)(weighted)= 2.37 into language redundancy(R) formula, we get the value of language redundancy
= 1 - 0.895 / 2.370 …. Eq. (5)
= 0.623 or 62.3%

H(X)(max)(weighted), and redundancy R would, however, change if the number of vowels and consonants are rounded off to 2 and 4 respectively for the mean word length of 6 letters.

H(X)(max)(weighted) = [1.660*2 + 2.696*4] / 6 …. Eq. (6)
= 2.35

By substituting H(X) = 0.916 (see eq. 1 (b)), and
H(X)(max) (weighted) = 2.35 into the formula for language redundancy (R), we get the value of language redundancy:
= 1 - 0.916 / 2.35 ……Eq. (7)
= 0.611 or 61.1 %

Bibliography

1. Robert G. Gallagher. Claude E, Shannon: A Retrospective on his Life, Work and Impact, IEEE Transactions on Information Theory, 47(7). 2681-2695. 2001. https://mast.queesu.ca.>gallager-on-shannon; https://www.researchgate.net>
2. Robert M. Gray. The Entropy and Information Theory. 61-95, New York: Springer-Verlag. 2013
3. E.T. Jaynes. Information theory and statistical Mechanics. Phys. Rev. 106(4) 620 -63.1983
4. Robert P.H. Gasser, W. Graham Richards. An Introduction to Statistical Thermodynamics, UK: World Scientific.1995
5. Atkins Peter. The Laws of Thermodynamics, Oxford, UK. Oxford University Press.2010
6. R.K. Pathria. Statistical Mechanics. Second edn. , Oxford ,UK. Butterworth -Heinemann.2017
7. E.T. Jaynes. Gibbs vs Boltzmann Entropies. American Journal of Physics,33, 391-398.1965
8. Alberto Gianinetti. An Account of Thermodynamic Entropy. Bentham Science Publisher.2017
9. C.E. Shannon. " A Mathematical Theory of Communication", (part 1). reprinted from The Bell System Tech. J., 27 (3), 379 -423. July 1948; (part 2), The Bell Lab System Tech. journal, 27 (4). 623-656. October 1948. https://www.people.math.harvard.edu
10. C.E. Shannon. Prediction and Entropy of Printed English, Bell Syst. Tech. J., 30, 47-51, Wiley Online Library. Jan1951.https://www.princeton.edu>shannon_51(1)(3)

11. Sebastian Kwiatkowski. Entropy is a measure of Uncertainty. 2018. Retrieved from https//www.towardsdatascience.com>entropy
12. Thomas. M Cover, Joy A Thomas. Elements of Information Theory. (2nd ed.) Wiley – Interscience.2006.
13. N.G. Burton, J.C.R. Licklider. Long-range constraints in the statistical structure of printed English. Amer. J. Psych., 68, 650-653.1955
14. James L. Dolby, Howard L. Resnikoff. On the Structure of written English words. Language 40 (2), 167-196.1964
15. J. Pierre Humbert, July. Phonological and phonetic representation, Journal of Phonetics, 18 (3), 375-394.1990
16. Ryan T. Miller. English Orthography and Reading. The TESOL Encyclopaedia of English Language Teaching. Wiley online Library.2019
17. V. V. Bochkarev, A.V. Shevlyakova, V.D. Solovyev. The average word length dynamics as an indicator of cultural changes in the society. Social Evolution and History. 14(2): 153-175. Uchitel Publisher. Volgograd, Russia.2015
18. Mark Davies . Examining syntactic variation in English: The importance of Corpus design and Corpus size, English Language and Linguistics. 19 (3):1-35.2014
19. Leo Qin .2018. https://www.leozqin.me>vowels-compressibility
20. Stanford education ,1990. retrieved from www.cs.stanford.edu/people/projects/1999-00/Information Theory/Entropy of English).
21. Hamid Moradi, J.W. Gryzymala-Bussee, James A. Roberts. Entropy of English text: Experiments with humans and machine learning system based on rough sets Information Science, An International Journal, 104 (1-2), 31-47.1998
22. A. Lempel, J. Ziv. Compression of individual sequences via variable-rate coding. IEEE Transactions on Information Theory, 24(5), 530-536. 1978.
23. T. A. Welch. A technique for high-performance data compression. IEEE Computer, 17(6), 8-19.1984
24. B. Mandelbrot. An Information Theory of the Statistical Structure of a Language in Communication Theory, Edited by W. Jackson, 486-502, New York: Academic Press.1953
25. D. Sperber, D. Wilson. Relevance: Communication and Cognition. (2nd Edn.) Oxford, U.K. Blackwell Publisher Ltd.1995
26. D. Wilson, D. Sperber. Linguistic form and relevance. Lingua. 90:1-25.1993
27. L. Cortellazzo, E. Bruni, R. Zampieri. A role of leadership in digitalised world. Front. Psychol., Sec. Organizational Psychology (10).2019
28. Robert B. Kaiser. The Best Leaders are Versatile ones. 2020. www.hbr.org/2020/03
29. Teresina Torre , Daria Sarti . The Way toward e-leadership: Some evidence from the field. Frontiers in Psychology. (11), Nov.2020. DoI:10.3389/fpsy.g.2020.554253; https://www.researchgate.net
30. Kim Girard. It is not nagging: why persistent redundant communication works. HBS News, 2011. https://hbswk.hbs.edu>item.its-