ORIGINAL PAPER

The Principle of Least Effort in Technical Register.
A Case Study on Students in Automation

Adrian-Florin Bușu[1]

Abstract
Also known as Zipf's Law, the Principle of Least Effort is the theory according to which one single primary principle in any human action, including verbal communication, represents the expenditure of the least ammount of effort to accomplish a task. When this principle is applied to linguistics, it can be reduced to the following observation: the more frequent a word is, the shorter it tends to be. Zipf claimed that this specific attribute occurs as a consequence of human attempt to maximize word connotation framings in circumstances dictated by factors such as dynamic constraints to communicate in an accurate manner, or the desire to efficiently exchange information. This article presents the results of a case study conducted for a period of 6 weeks on students in the first year at the Faculty of Automation, Computers and Electronics from University of Craiova and provides direct experimental evidence to support Zipf's explanatory hypothesis. The findings of this case study suggest that the communicative pressure acting on language users is directly responsible for the frequency relationships in the vocabulary. This article promotes an approach on language as an efficient code for information transfer and concludes that lexicon naturally evolves towards greater efficiency.

---

[1] Lecturer, PhD, University of Craiova, Faculty of Letters, Department of Applied Modern Languages, Romania, Phone 0040744177449, Email: adibusu2002@yahoo.com, ORCID ID 0000-0003-1379-9918.

I. Introduction

In "L'inertie mentale et la loi du moindre effort", published in "Revue Philosophique de la France et de l'Étranger" in 1894, Guillaume Ferrero presented the fundamentals of the principle of achieving more by doing less. Naturally, such an appealing approach on the topic of least effort stirred the interest of some experimentalists, who conducted a series of scientifical experiments to support this theory. In 1930, J. A. Gengerelli published the results of several experiments conducted on blinded and normal albino rats. The purpose of the experiments was to identify the type of path the subjects would choose from a number of possible routes leading to food. What is interesting though, is that the results clearly indicated that the path chosen by both categories of rodents was the one of least effort.

In this respect, according to Tsai, who claimed that "Among several alternatives of behavior leading to equivalent satisfaction of some potent organic need, the animal, within the limits of its discriminative ability, tends finally to select that which involves the least expenditure of energy" (Tsai, 1932: 18), we gather that the principle of least effort applies not only to humans, but, to a certain extent, even to animals. Waters connected the results of the above-mentioned experiment to the epic character of Theseus, who managed to slay the the minotaur and, after that, to figure out how to get out of the labyrinth, by following the string on his way out: "Perhaps, this was not the most direct route in terms of distance, time, or effort, but it was the only sure way he had of escaping. Likewise our rats found that by sticking to the outside pathways, they more readily achieved the goal". (Waters, 1937: 51)

"Each individual will adopt a course of action that will involve the expenditure of the probably least average of his work". This assertion was entitled "The Principle of Least Effort" and is also known as "Zipf's Law", "Zipf's Principle of Least Effort" or "The Path of Least Resistance". This principle was formulated in 1949 in "Human Behaviour and the Principle of Least Effort", a book written by G. K. Zipf, professor of Philology at Harvard University. In his book, Zipf hypothesized about the correlation between the length and the frequency of word and he theorized that the more frequent a word is, the shorter it tends to be.

The researcher stated that the tendency to communicate efficiently with minimal effort generates the distribution of words in communication. Based on this principle, Zipf went on to generalize that it is the human nature itself to pursue the maximal outcome at the minimal amount of effort. Zipf claimed that this Law of Abbreviation is a universal structural property of language. The Law of Abbreviation has been extended to domains such as animal communication systems or computer programming languages. Zipf claimed that this specific characteristic occurs as a consequence of human attempt to maximize word connotation framings in circumstances dictated by factors such as dynamic constraints to communicate in an accurate manner, or the desire to efficiently exchange information. As a result of this phenomenon, it is suggested that reccurent types of behaviour become quicker and easier to perform over time and, in many cases, people customize their behaviour along the direction of minimizing the effort.

II. Problem statement
1. Zipf's Principle of Least Effort related to people's social behavior

In an attempt to connect the Principle of Least Effort to the social behaviour, Zipf theorized that the individual will try to find a solution to a problem in a manner

that would bring to a minimum the ammount of effort expended in that regard: "In simple terms, the Principle of Least Effort means, for example, that a person in solving his immediate problems will view these against the background of his future problems, as estimated by himself. Moreover, he will strive to solve his problems in such a way as to minimize the total work that he must expend in solving both his immediate problems and his probable future problems". (Zipf, 1949: 63) Having set this definition as a starting point, Zipf proceeded to analyse the least effort of individual and collective behaviour from two distinctive approaches. In the case of individual behaviour, he performed an analysis on specific structures of children's communication, whereas in the case of collective behaviour, he concentrated his attention on more general aspects such as human social behaviour and the distribution of social status. Zipf started from the experimental aspect, making a series of interesting observations from an extensive range of angles. He presented the fundamental theoretical analysis and synthesised various types of experimental principles in a unique general law. As a result of Zipf's perspective and the fundamentals of his law, the Principle of Least Effort can be interpreted from both experimental and theoretical approaches. Zipf's area of interest was the statistical study of the frequency of word use, but his principle has also been applied in Linguistics to topics such as lexical diffusion, language acquisition and conversation analysis. In addition to this, the Principle of Least Effort has been used in a series of other disciplines, such as Psychology, Sociology, Economics, Marketing and Information Science.

2. Language Changes and the Principle of Least Effort

The phenomenon by which permanent alterations and modifications are made in the features and the use of a language over time is known as language change. All languages are subject to language change and this phenomenon affects all areas of language use. There are several explanations for language changes, such as migration, technology, analogy or imperfect learning. Millward suggested that the Principle of Least Effort tends to result in phonetic reduction of speech forms, thus leading to language changes. He went on to explain that speakers, regardless of their educational background, unconsciously tend to simplify their speech, in some cases because they are simply careless, in other cases because of the pressure to communicate the maximum ammount of information at the expense of the minimun ammount of effort. Millward offers some examples in this regard, indicating that people use abbreviated forms like *Math* for *Mathematics* and *plane* for *airplane*, whereas *going to* becomes *gonna*, most frequently in spoken English, stressing the fact that the contracted forms have less phonemes to articulate. Moreover, Millward discusses the changes that occur on the morphological level, showing that speakers tend to make use, in some cases, of the past tense form of the verb instead of third form of irregular verbs, such as *showed* or *spoke* instead of *shown* or *spoken*, as there will be one less irregular verb form to remember for them. He concludes that The Principle of Least Effort might be a satisfactory explanation for many changes in language and "probably plays an important role in most systemic changes, such as the loss of inflections in English". (Millward, 1996: 86)

3. Writing Systems and the Principle of Least Effort

Writing systems are affected by the various changes that may occur at all language levels and which are determined by the Principle of Least Effort. Coulmas

supports this theory and claims that "the inventory of basic signs is small and can be easily learned, whereas it asks for substantial efforts to master a system with an inventory of thousands of elementary signs, like the Sumerian or Egyptian, which did what the Chinese, according to the evolutionary theory, should have done, namely give way to a system which can be handled with greater ease. This kind of thinking is reminiscent of Zipf's Principle of Least Effort". (Coulmas, 1991: 41) In other words, from the Proto-Sinaitic script to the Phoenician, Greek and Latin alphabet, there were a multitude of variants that evolved as a result the changes in the language caused by the speakers' attempt to overcome phonological limitations in communication or by their desire to minimize articulatory effort.

4. Applications of Zipf's Principle

In "Foundations of Statistical Natural Language Processing", Manning and Schütze supported Zipf's principle, claiming that "Zipf's law is useful as a rough description of the frequency distribution of words in human languages: there are a few very common words, a middling number of medium frequency words, and many low frequency words. Zipf saw in this a deep significance" (Manning & Schütze, 1999: 90). This statement supports the hypothesis according to which in the communicative flux, both the speaker and the hearer tend to minimize their effort by using reduced lexicon, contractions or less phonemes. Despite these changes in the communicational pattern, the meaning does not become ambiguous, as the linguisitc compromise does not affect the message to be transmitted. Case indicated further applications of Zipf's Law in modern domains: "The Principle of Least Effort has been most recently applied as an explanation in the use of electronic resources, most notably Web sites" (Case, 2001: 104). This observation indicates that in the near future, Zipf's Principle may affect other domains, as communication is not restricted only to face-to-face interaction. Newsgroups on the Internet, for instance, may become environments in which the Principle of Least Effort might generate changes, just as in real-life spoken languages (Bărbuceanu, 2020: 137).

5. Types of distribution. The Probability, Pareto and Zipf distributions

In the domain of experimental analysis, the specialized literature concluded that there are three main conventional distributions: the Probability Distribution, the Pareto Distribution and the Zipf Distribution. A series of natural phenomena, such as volcanic eruptions, tornadoes paths and earthquake magnitudes, occurs following the pattern of the Probability Distribution. The Zipf Distribution concentrates on the interconnection between the frequency of an event and the category to which the event belongs. The Pareto Distribution is a progressive distribution, illustrating that many things are not distributed evenly. Although these three types of distribution refer to the same concept, but from different perspectives, the Zipfian distribution seems to have a larger degree of relevance for artificially and naturally occuring phenomena.

6. The impact of Zipf's Principle on the field of communication

There is a predilection in the field of communication towards efficiency and economy, which consists in using compressed forms such as abbreviations of words and phrases, contractions or acronyms, based on the relationship between the overall incidence of a word, a phrase or a sentence and its size. To put it another way, scientists and researchers have always used abbreviations and acronyms of various types to

facilitate rapid and precise technical communication. Abbreviations such as *etc.*, *et al.*, *op. cit.* and *N.B.* are still part of the academic discourse. Similar simplistic tendencies exist in everyday discourse and especially in internet-based communication, such as emails and text messages. Various theories have been proposed in order to explain why languages change. One of the most interesting theory was stated in the 1950's by André Martinet, a French linguist. He claimed that languages change as a result of the operation of economic tendencies. Calling it "The Principle of Economic Change", Martinet pointed out that complex language forms and structures tend towards reduction, abbreviation, compression or elimination over time. For example, the contrast between short and long vowels in Latin, which generated a significantly substantial set of distinct words in Latin, was levelled in the emerging sound systems of the Romance languages and then later removed. The phonetic system in Latin had ten distinct vowel sounds, equivalent to the vowel phones represented by the letters *a*, *e*, *i*, *o* and *u*. Furhermore, each vowel was spelt as either long or short vowel, therefore words could have more than one meaning, depending on whether the vowel was pronounced long or short. The ten-vowel phoneme system was, to a large extent, simplified in the Romance languages, in accordance with the Principle of Economic Change. Distinctions of meaning were maintained, but with fewer phonemes. Grammar plays an important role in this respect. In English, for instance, the first word in a sentence is a noun which acts as the subject that performs the action, or the doer. The mere fact that the first word is a noun imposes limits on the chances for the next word to be the subgroup of verbal phrases. Depending on the type of verb chosen, the following words function as object and are yet again more limited. It is not just the grammatical structure that enforces successive restrictions; the need for intelligibility has a similar effect. The lack of hierarchical structures in the formation of sentences causes a lax interpretation of their meaning. On specific occasions and in certain circumstances, words play the role of a linguistic joint, meaning that it allows for more successive words than were available for its preceding word. Faulty articulation of sentences triggers a certain degree of ambiguity in the linguistic code. For language in particular, any such account of Zipf's principle provides a psychological theory about what occurs at the comprehensive and interpretative levels of language users.

7. Further insights into Zipf's Principle of Least Effort

This section reviews a wide range of theories on word frequency distributions, but purposefully avoids statistical evidence as this is not the focal point of this article. Instead, we focus here on facts about word frequencies that are informative about the mechanisms which enforce Zipf's principle. One question arises though: how do the complex processes of normal human language production work together in order to result in a frequency distribution that is so simple from a mathematical perspective? This question has represented a main area of interest for researchers in the field of statistical language theories over the past 70 years. There are numerous derivations of Zipf's principle from more fundamental hypotheses, both in language and in other fields of science where this law can be encountered. (Edmunson, 1972) Explanations for the distribution across various fields of science include many frameworks and sets of assumptions. To have a clear image of the range of explanations that have been offered, such distributions have been claimed to originate in random concatenative processes, mixtures of exponential distributions, bounded optimization of entropy, Fisher information, the invariance of power laws under aggregation, multiplicative stochastic

processes, preferential reuse, symbolic descriptions of complex stochastic systems, semantic organization, communicative optimization, random division of elements into groups, first and second order approximation of most common distributions and optimized memory search. (Edmunson, 1972: 84) Most papers on language research have focused mainly on deriving the law itself in principle and just a few have evaluated the fundamental suppositions of the hypothesized explanation. (Rossen & Resnick, 1982: 44)

III. Research questions

The aim of the study is to present the characteristics of empirical phenomena and to use them to create specific model-based accounts of Zipf's law, starting from the observation that the distribution of word frequencies is more intricate than expected.

An analysis of the importance of the best distributional form is not the focal point of this section, as what really matters are the general properties of word frequency that are explanatory about the fundamental mechanisms of the analysed distribution. Instead, the present paper is an attempt to offer answers to some questions: To what extent does the Zipfian distribution occur for specialized terminology in the technical register of the language? Is this type of distribution encountered with technical words as in, for example, English for Information Technology, or is it restricted to common words? To answer these questions, in this study we have analysed relevant subgroups of the lexicon, namely the technical register of language.

IV. Purpose of the study

The study was conducted for a period of 6 weeks on a group of 50 students, first semester, freshman year, from the Faculty of Automation, Computers and Electronics, University of Craiova. The purpose of the study was to identify which are the most frequently used words both in regular and technical register of the language, as well as to identify to what degree the Principle of Least Effort affects the communication among students, if it causes changes in language and if it optimizes information exchange.

V. Research methods

The methods used for this study were survey, participant observation and secondary analysis. The frequency of words was measured by occurence per 50000 words in speaking and writing, while the degree of efficiency was measured by observing the level of comprehension.

VI. Findings

The present paper provides experimental evidence that the Zipfian word frequency distribution occurs for both common and technical words in a language production task. After analyzing the results of the survey, we can conclude that the most used three words were *the*, *of* and *and*, as expected. No surprise here, as the results confirm Zipf's hypothesis and other similar surveys and studies, for example an analysis performed on *The Declaration of Independence*:
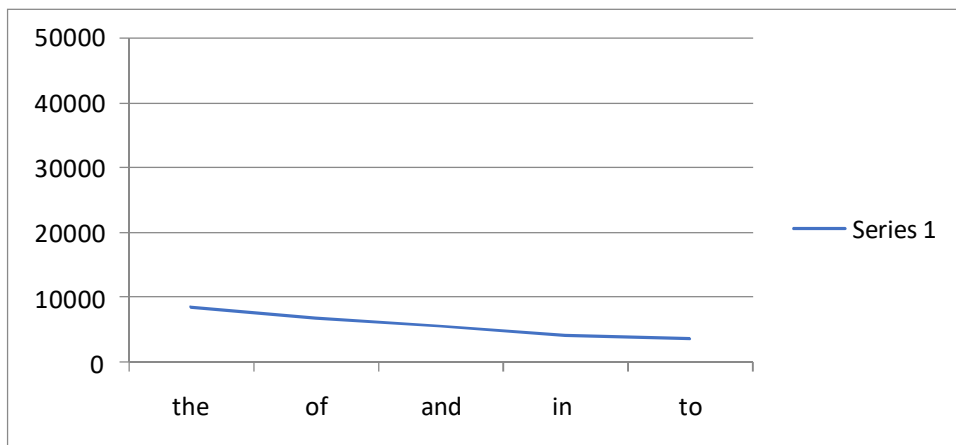
Table 1. Analysis of word frequency in *The Declaration of Independence*

Source: https://demonstrations.wolfram.com/ZipfsLawForNaturalAndRandomTexts/
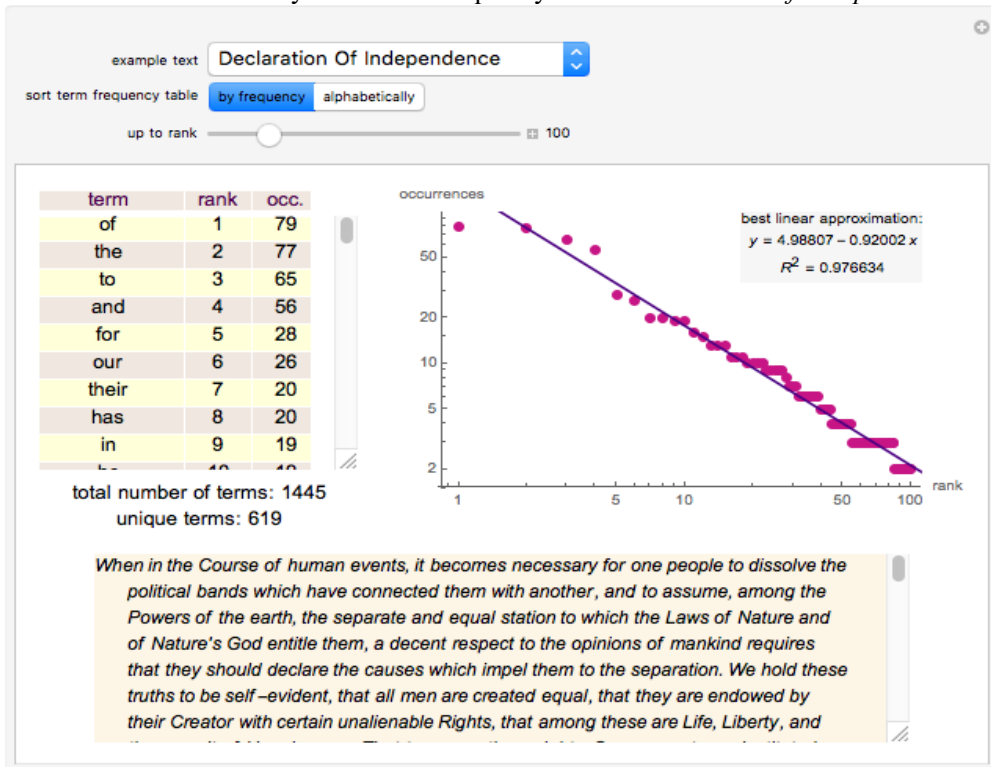
Table 2. Results of survey - word frequency_length

However, one really interesting finding is that the most used three technical words were *RAM*, *CPU* and *CD*, which are actually abbreviations for *Random Access Memory*, *Central Processing Unit* and *Compact Disc*. This demonstates that abbreviation is a means of optimizing technical communication that students rely on.

Table 3. Results of survey - word frequency_technical words
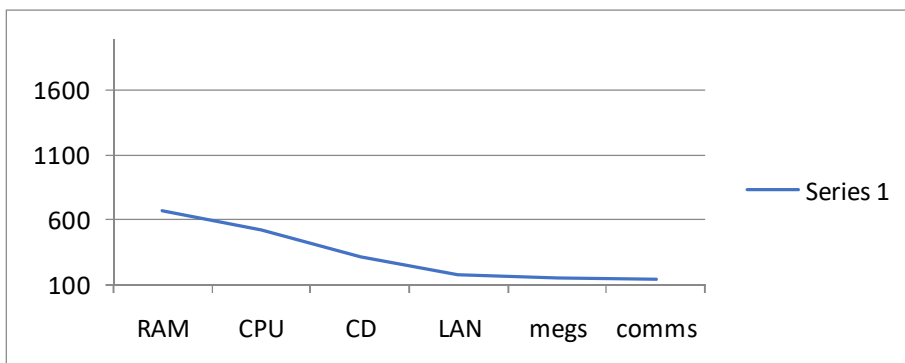
Another interesting finding of this survey is that students tend to maximize communication by reducing complex grammatical structures to shorter, less cumbersome lexical constructions, unwittingly using compression, reduction or even or elimination of morphemes. Following this observation and the results of the survey, we can note that students, for example in *I haven't spoke with the sysop in a long time*, tend to replace the past participle of the verb in the present perfect with the past tense form of the verb, by eliminating the final consonant of the past participle, thus forming a structure that in theory shoud be avoided, but which yet is still comprehensible. A quick search on *textranch.com* reveals a whooping 41900 indexed results of this hybrid grammatical structure, meaning that such a type of language change has made its way into communication.

Table 4. Results of survey – word frequency_reduction

Source: www.textranch.com/370043/i-havent-spoken/or/i-havent-spoke/

VII. Conclusions

All things considered, we believe that, in light of the aspects presented in this article and the results of the survey, it is worth reflecting on the distinctiveness of Zipf's principle. The evidence that words vary in frequency is beyond any doubt and represents an important feature of language. It is not quite clear why words vary in frequency, as it is difficult to pinpoint the reason for this. What is certain is that we produce speech because we need to communicate a meaning in a specific social or cultural context and our statements comply with complex rules at syntactic, lexical and semantic levels. However, we unconsciously tend to adjust the parameters of our communication along the direction of minimizing the effort. The findings of this case study suggest that the communicative pressure acting on language users (students in Automation, Computers and Electronics, in our case) is directly responsible for frequency relationships in the vocabulary. Taking into account the features of information transfer as presented in this article, we can conclude that lexicon naturally evolves towards greater efficiency, especially in the technical register of the language.

References:

Bărbuceanu, D. (2020). *Teaching the Digital Natives*, in *Revista de Ştiinţe Politice*, no. 65. Universitaria Publishing House

Case, D. O. (2001). *The Principle of Least Effort. Theories of Information Behavior*, edited by Karen E. Fisher, Sandra Erdelez and Lynne [E.F.] McKechnie, Information Today

Coulmas, F. (1991). *The Influence of Language on Culture and Thought: Essays in Honor of Joshua A. Fishman's Sixty-Fifth Birthday*, ed. by Robert L. Cooper and Bernard Spolsky. Walter de Gruyter

Edmundson, H. P. (1972). *The Rank Hypothesis: A Statistical Relation between Rank and Frequency Technical report*. TR-186

Ferrero, G. (1894). *L'inertie mentale et la loi du moindre effort*", in *Revue Philosophique de la France et de l'Étranger*, 37, 169-182

Gengerelli, J. A. (1930). *An experiment on the natural preference of rats for right and left turns* in *Journal of Comparative Psychology, 10*(3)

Manning, C. D., Schütze, H. (1999). *Foundations of Statistical Natural Language Processing*. The MIT Press

Martinet, A. (1975). *Évolution des langues et reconstruction*, Paris: PUF

Millward, C.M. (1996). *A Biography of the English Language*, 2nd edition, Harcourt Brace

Rossen K. T., Resnick M. J. (1982). *Urban Economy*. No/8, pp. 165-186

Tsai, T. S. (1932). *The laws of Minimum Effort and Maximum satisfaction* in *Animal Behavior Monograph of the National Institute of Psychology*

Waters, R. H. (1937). *The Journal of General Psychology*, Taylor & Francis Publishing House

Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*, Addison-Wesley Press.

Adrian-Florin Bușu