Zeitschrift: Studies in Communication Sciences : journal of the Swiss Association of Communication and Media Research

Herausgeber: Swiss Association of Communication and Media Research; Università della Svizzera italiana, Faculty of Communication Sciences

Band: 6 (2006)

Heft: 1

Artikel: Alphabets and the principle of least effort

Autor: Danesi, Marcel

DOI: https://doi.org/10.5169/seals-791093

Studies in Communication Sciences 6/1 (2006) 47-62

MARCEL DANESI*

ALPHABETS AND THE PRINCIPLE OF LEAST EFFORT

Alphabet systems have made the recording of information an efficient matter. As a consequence, they have made it possible for human civilizations to progress quickly and expansively. Alphabet characters are derivatives of pictographs, allowing for a more condensed means of recording and transmitting knowledge. The purpose of this paper is to argue that alphabets came about, in fact, to do just this — namely, to make knowledge representation efficient. One of the first to study the "efficient" nature of letters empirically was the Harvard linguist George Kingsley Zipf, who demonstrated that there is universally a correlation between the length of a specific word (in number of letters) and its rank order in a language. This paper will look at Zipf's work and assess its importance to semiotic theory, especially as it relates to the nature of signs and how they express meaning.

Keywords: alphabets, semiotic theory, communication theory, Zipf's Law, principle of least effort, quantitative linguistics.

* University of Toronto, marcel.danesi@utoronto.ca

Introduction

The development of alphabets over 3000 years ago has been proclaimed by communication scientists (and other scholars) as one of the most significant events in human history after the invention of tools. Before that, people transmitted knowledge primarily through the spoken word, even though pictography was used to record and preserve ideas in some more durable way. Pictography is so basic and instinctive that, even in alphabet-using cultures, it has hardly disappeared. It can, in fact, be seen all over the modern cultural landscape, from figures designating *male* and *female* on washrooms, to logos and trademarks in advertising. Alphabets are extensions of pictographic sign systems. They emerged, arguably, to allow for a more efficient means of recording and transmitting knowledge, by providing symbols to represent phonemes or other sound structures, rather than concepts (as did pictographs). Known as the Alphabet Principle, the symbol-for-sound basis of alphabetic writing has not only altered the ways in which knowledge is transmitted, but also how we have come to perceive it and even give it form. In his main writings, Marshall McLuhan (1951, 1958, 1964) asserted that in allowing for efficient knowledge storage and retrieval this Principle came forward to trigger the first true "cognitive revolution" of humankind. The reason for this, as Eco (1994: 167) aptly pointed out, was that the "alphabet did more than represent words, it analysed them as well."

The importance of the Alphabet Principle to the constitution and organization of civilization has been discussed abundantly and repeatedly in many fields (ranging from linguistics and semiotics to philosophy and archeology), as has its emergence to replace pictography as an efficient method for recording knowledge (e.g. Gelb 1963; Arnheim 1969; Ong 1977; Posner 1983; Gaur 1984; Harris 1986; Logan 1987; Coulmas 1989, 2003; Watt 1994; Daniels & Bright 1995; Drucker 1995; Martin 1999; Fischer 2001). But, to the best of my knowledge, rarely (if ever) has anyone contemplated ascertaining why such a replacement came about in the first place. It is an implicit assumption in most treatments, not an explication. The general "storyline" offered in the scientific literature is that alphabet characters, as stylized derivatives of pictographs, came to replace the latter in the marketplaces of the ancient world when people started realizing that they could be used efficiently (over and over) to represent the sounds in words. That watershed event in human history is thus typically portrayed as a fortuitous social episode, not as a product of some psychobiological "propensity" at work in the human species. Simply asserting that alphabets came about to render the encoding and use of information efficient does not in any way explain why efficiency is a propensity in representation and communication.

As is fairly well known in both linguistics and semiotics, one of the first to study this "propensity" empirically in the 1930s and 1940s was the Harvard linguist George Kingsley Zipf (1902-1950). Essentially, Zipf claimed that many phenomena in language could be explained as the result of an inborn tendency in the human species to make the most of its signifying resources with the least expenditure of effort (physical, cognitive, and social). Fortunately, his claim did not end up in the usual scrap-heap of speculation that tends to characterize theorizing of this sort, for the simple reason that he applied it as a "litmus test" to examine any possible relation between the physical nature of linguistic forms and their frequency of occurrence empirically, discovering serendipitously that the relation could be captured by a mathematical formula (Zipf 1935, 1949). Known as Zipf's Law, much research has been conducted in linguistics and other sciences to validate or refute it, with the constant outcome being its validation. Basically, Zipf demonstrated that there is an intrinsic interdependence between the length of a specific word (in number of phonemes) and its rank order in the language (its position in order of its frequency of occurrence in texts of all kinds). The higher the rank order of a word (the more frequent it is in actual usage), the more it tends to be "shorter in length" (made up with fewer phonemes). For example, articles *(a, the)*, conjunctions *(and, or)*, and other function words *(to, it)*, which have a high rank order in English (and in any other language for that matter), are typically monosyllabic, consisting of 1-3 phonemes. What is even more intriguing is that this "compression" force does not stop at the level of function words, as Zipf and others subsequently found. It can also be seen to manifest itself, for instance, in the tendency for phrases that come into popular use to become abbreviated in some way (*FYO, UNESCO, Hi, Bye, 24/7,* etc.). In effect, the general version of Zipf's Law proclaims that the more frequent or necessary a form for communicative purposes, the more likely it is to be rendered "compressed" or "economical" in physical structure. And the reason for this seems to be an inherent psychobiological tendency in the human species to expend the least effort in representation and communication.

In semiotic terms, Zipf's Law can be formulated more precisely as the tendency to reduce or compress the signifiers in a code, while preserving the signifieds. In effect, the Law implies that the meaning relation between signifiers (forms) and signifieds (referents) is preserved, even though the number and size of the signifiers in a code tend towards diminution.

The purpose of this paper is to revisit the Alphabet Principle from the perspective of this general version of Zipf's Law which, surprisingly, has received relatively little attention within semiotics proper. It is important to note that this is not a quantitative study of the Law as applied to alphabets among other symbol systems. There is plenty of literature on that aspect of the Law. The goal here is a more generic implicational one—namely that it will allow semiotics to get a better grasp of why signs and sign systems tend toward structural economy. Indeed, future work in semiotics and communication theory would benefit greatly by using Zipf's insights, which arguably provide a framework for studying the passage from iconic to symbolic representation empirically, not just speculatively (as is often the case). A "Zipfian approach" in semiotics and communication theory would make it possible to entertain the more general question of whether symbolism is result of the tendency towards least effort and, if so, why it appears to be unique to the human species (Sebeok 2001; Sebeok & Danesi 2000).

Zipf's Law and The Principle of Least Effort

Since the mid-1950s, Zipfian-inspired research has established empirically that there is a tendency in human sign systems towards the compression of high-frequency forms (Miller & Newman 1958; Kucera & Francis 1967; Wyllys 1975; Rousseau & Zhang 1992; Li 1992; Ridley & Gonzales 1994; Perline 1996; Nowak 2000). This tendency can be seen, for example, in the use (and constant growth) of abbreviations and acronyms: *ad=advertisement; NATO=North Atlantic Treaty Organization, laser=light amplification by stimulated emission of radiation*, etc. It can also be seen in tables, technical and scientific notation systems, indexes, footnotes, bibliographic traditions, and so on and so forth. All such economical phenomena validate the underlying hypothesis in Zipf's Law that compression saves effort. Remarkably, Zipf's Law has been found to characterize many types of human representational activities and behaviors, from numeration patterns (Raimi 1969) to the distribution of city populations (Burke & Kincannon 1991; Hill 1998).

To grasp the essence of the Law, all one has to do is take all the words in a substantial corpus of text, such as an issue of the *New York Times*, counting the number of times each word in it appears. If the frequencies of the words are then plotted on a histogram and sorted by rank, with the most frequently appearing words *(a, the, for, by, and)* first, then the resulting curve will be found to approach the shape of straight line with a slope of -1. A study of Zipfian histograms has revealed some truly remarkable tendencies: (1) the magnitude of words tends, on the whole, to stand in an inverse relationship to the number of occurrences (the more frequent the word the shorter it tends to be); and (2) the number of different words in a text seems to be ever larger as the frequency of occurrences becomes ever smaller. In the figure below (adapted from Cherry 1957: 104-106), curve **A** shows the result of a word count made upon James Joyce's *Ulysses*, which contains nearly 250,000 word tokens with a vocabulary of nearly 30,000 word types:

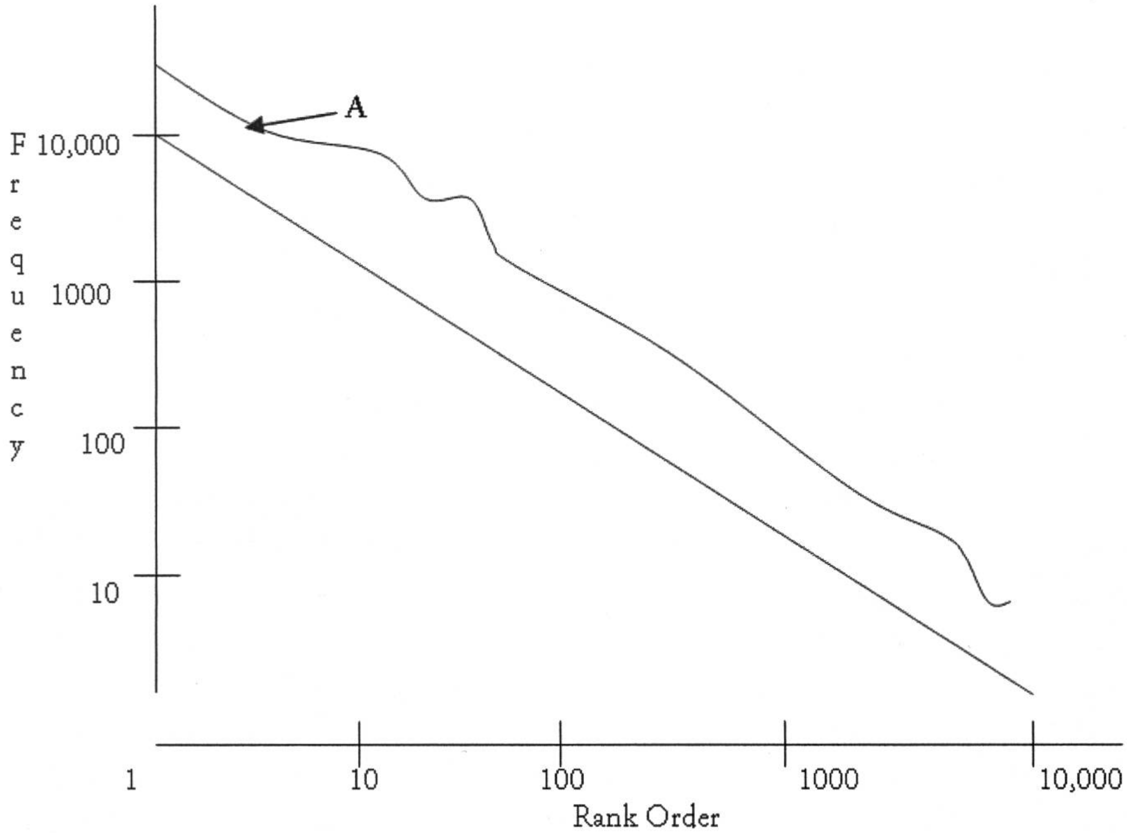

Figure 1: Frequency Graph of James Joyce's "Ulysses"

The curve represents the frequency of the word-types in Joyce's novel against rank order. The slope is downward from left to right, approaching the value of –1 (the straight line in the middle). This result seems to emerge no matter what type of text is used. Indeed, given a large enough corpus, the exact same type of curve results from counting words in newspapers, textbooks, recipe collections, and the like. The larger the corpus the more the curve tends towards the slope -1. The language also does not influence this result. Indeed, Zipf used data from widely-divergent languages and found this to be true of any language. Not only words, but phonemes, syllables, morphemes, Chinese characters, and the babbling of babies have been found to fit the Zipfian curve.

Given this constant outcome, the relation of word (morpheme) frequency (p_n) to rank order *(n)* was formalized by Zipf as follows:

$$\log p_n = A - B \log_n \text{ (where A and B are constants and } B \approx 1)$$

For the sake of historical accuracy, it should be mentioned that this type of outcome was known long before Zipf. Already in the nineteenth century, it was found that if the digits used for a task (to enumerate, classify, etc.) are not entirely random but somehow socially or naturally based, the distribution of the first digit is not uniform — 1 tends to be the first digit in about 30% of cases, 2 will come up in about 18% of cases, 3 in 12%, 4 in 9%, 5 in 8%, etc. This was discovered by the American astronomer Simon Newcomb in 1881, who noticed that the first pages of books of logarithms were soiled much more than the remaining pages. In 1938, mathematician Frank Benford investigated listings of data, finding a similar pattern to that uncovered by Newcomb in income tax and population figures, as well as in the distribution of street addresses of people listed in phone books. Zipf's main contribution was in showing empirically that patterns of this type manifest themselves regularly and almost "blindly" in human representational efforts, especially in language.

Shortly after the publication of Zipf's Law, the mathematician Benoit Mandelbrot (1924-), who developed the modern-day branch of mathematics known as fractal geometry, became fascinated by it (Mandelbrot 1954, 1983), since he detected it as being a particular type of what is called a "scaling" law in biology. As a brilliant mathematician, Mandelbrot also made appropriate modifications to Zipf's original Law and, generally speaking, it is Mandelbrot's version of the Law that is used today to study frequency distribution phenomena in several branches of linguistics, such as corpus linguistics, lexicostatistics, glottometrics, textlinguistics, and quantitative linguistics generally.

The most fascinating aspect of Zipfian analysis is that it suggests the unconscious operation of a Principle of Least Effort in the human species, which, as Zipf himself claimed, governed our individual and collective representational and collective behavior. Clearly, this Principle (if valid) is pivotal for understanding the emergence and diffusion of alphabets. The Principle of Least Effort can be reformulated in semiotic terms as the Signifier Compression Principle (SCP), which states explicitly that the tendency in representation and communication systems is towards the compression of the signifiers at the same time that range of the signifieds in the system is preserved. The SCP is, in some cases, related to time. The longer words and expressions have been around, for instance, the more likely they will be abbreviated (*etc.* = *et cetera, N. B.* = *nota bene, Q. E. D.* = *quod erat demonstrandum, et. al.* = *et alibi,* and so on) for the simple reason that they are used frequently. However, this is not always the case. The abbreviation *PC* = personal computer was compressed almost the instant it entered into currency. The reason in this case is that it referred to something that was used constantly. Simply put, compression is the outcome of the tendency to reduce effort, either because of usage over time or because of immediacy of utilization.

The SCP operates at all levels and dimensions of language. It can be seen, for example, as a factor in sound shift. In the Romance languages, the outcome of the Latin cluster /kt/, for instance, reflects an attempt to mitigate the gap between the /k/ sound, which is articulated in the back of the throat, and the /t/ sound which is articulated at the front end of the mouth. Phonetically, the distance between these two sounds makes it effortful to articulate the cluster /kt/, as readers can confirm for themselves by pronouncing the Latin words *noctem* ("night"), *octo* ("eight"), and *tectum* ("roof") slowly. Assimilation — the process whereby one sound takes on the characteristic sound properties of another, either partially or totally — has intervened to make the articulation much more effortless, by either gapping the distance between /k/ and /t/ or eliminating it altogether. The Italian outcomes *notte, otto,* and *tetto*, for example, show complete assimilation. The outcome *ch* (= /ĉ /) in Spanish (*noche, ocho*, and *techo*) is a sound articulated midway between /k/ and /t/. Such outcomes of the cluster /kt/ reflect efforts to mitigate the gap between the /k/ and /t/ sounds. They are manifestations of the operation of the SCP.

The Emergence of Alphabets

As mentioned, writing started out as pictography. The *pictograph* is a signifier designed to simulate or illustrate some referent visually. The earliest pictographic instruments have been unearthed in western Asia. They are elemental shapes on clay tokens from the Neolithic era that were probably used as image-making moulds (Schmandt-Besserat 1978, 1992). One of the first civilizations to institutionalize pictographic writing as a means of recording ideas, keeping track of business transactions, and transmitting knowledge was the ancient Chinese one (Billeter 1990). According to some archeological estimates, Chinese pictography may date as far back as the fifteenth century BC. Pictographs that attempt to encode abstract ideas are called *ideographs*. These bear resemblance to their referents, but they entail conventionalized knowledge of the resemblance pattern on the part of the user. International symbols for such things as public telephones and washrooms today are examples of ideographs. Highly conventionalized ideographs, which tend to be combinations of pictographs, are called *logographs*. For example, the Chinese logograph for *east* is a combination of the pictographs for *sun* and *tree*.

A pictographic system was also used by the ancient Sumerian civilization that emerged nearly five thousand years ago. Called *cuneiform*, because it consisted of wedge-shaped picture symbols inscribed on clay tablets, it was an effective, but expensive and impracticable, means of written communication. For this reason, it was used primarily by rulers and clerics (Walker 1987). From about 2,700 to 2,500 BC a similar type of pictographic script, called *hieroglyphic,* was invented in Egypt (Davies 1988). This was somewhat more practicable than the Sumerian one, because the Egyptians used it not only for inscriptions on walls and tablets, but also for recording transactions and knowledge on papyrus (a much cheaper and portable medium). Incidentally, pictography was originally thought to have sacred or divine power, and this is why most early societies attributed the origin of writing to deities — the Cretans to Zeus, the Sumerians to Nabu, the Egyptians to Toth, the Greeks to Hermes, and so on.

Some pictographic systems (such as the hieroglyphic one) eventually developed *phonographic* elements within them — *phonographs* are signs standing for parts of words, such as syllables or individual sounds. The first true *syllabaries* — systems of phonographs representing syllables — were developed by the Semitic peoples of Palestine and Syria during the last half of the second millennium BC. Syllabaries are still used in some cultures today. Japanese, for example, is still written with two complete syllabaries — the *hiragana* and the *katakana* — devised to supplement the characters originally taken over from Chinese. A phonographic system for representing single sounds is called *alphabetic*. The first alphabetic system, which contained symbols for consonant sounds only, surfaced in the Middle East, and was transported by the Phoenicians (a people from a territory on the eastern coast of the Mediterranean, located largely in modern-day Lebanon) to Greece. The Greeks added symbols for vowel sounds, making their writing system the first full-fledged alphabetic one.

It is likely that iconic (visual-simulative) writing gave way to phonographic-symbolic (sound) writing in the marketplaces of the ancient world because, in many cases, it made the writing of transactions more rapid and efficient. Drawing an entire pictograph takes more time than writing it in outline form or in some abbreviated fashion. The transition was evolutionary, however, not revolutionary. Generally speaking, an alphabet character is the compressed residue of a stylistic alteration to some earlier pictograph. Take, for example, the emergence of the alphabet character **A**. It started out as a pictograph of the head of an ox in Egypt (and other parts of the Middle East). The full head of the ox came at some point to be drawn only in its bare outline, especially in Semitic cultures. It was this outline that came eventually to stand for the word for ox (*aleph* in Hebrew). Around 1000 BC Phoenician scribes, who wrote from right to left, drew the ox outline sideways (probably because it was quicker and "more natural" for them to do so). The slanted Phoenician figure came to stand just for the first sound in the word (*a* for *aleph*), because it became very familiar. The Greeks, who wrote from left to right, turned the Phoenician figure around the other way. Around 500 BC, as such "abbreviated picture writing" became more standardized and letters stopped changing directions, the *A* assumed the upright position it has today in Roman script — the ox had finally settled on its horns! (For a good summary of the development of writing from pictography to alphabets, see Driver 1976; Robinson 1995; Samoyault 1998).

When the Greeks started the practice of naming each symbol by such words as *alpha, beta, gamma*, etc., which were borrowings of Phoenician words (*aleph* "ox," *beth* "house," *gimel* "camel," etc.), it was evidence that the Alphabet Principle had crystallized completely in human consciousness. Incidentally, the order of the western alphabet (*A, B, C,...*) is derived from the fact that the sequence of letters from *A* to *Z* was used to count the numbers in order — *A* stood for 1, *B* for 2, *C* for 3, and so on. The earliest record of alphabetic order is, actually, Psalm thirty-seven where verses follow the Hebraic sequence, although, as Psychoyos (2005) has recently suggested, the double use of alphabets — for phonemic and numerical representation — was a widespread one in the Eastern Mediterranean and parts of Europe.

The Alphabet Principle

Pictography did not alter the basic oral nature of daily communication in early societies, nor did it alter the oral mode of transmitting knowledge. That occurred after the emergence of alphabetic writing — an event that brought about the first true revolution in how we record and understand knowledge. The move away from pictographic to alphabetic writing was, to use philosopher Thomas Kuhn's (1922-1996) appropriate term, one of the first great "paradigm shifts" in human cognitive and cultural evolution, constituting the initial event towards the establishment of a worldwide civilization (Kuhn 1962). The second shift occurred in the fifteenth century after the development of movable type technology — an event that made it possible to print and duplicate books cheaply. As a result, more books became available and more people aspired to gain literacy. With literacy came exposure to new ideas and independent thinking. And with independent thinking came many ideological revolutions of a religious, political, social, and scientific nature. Ideas started crossing borders and vast spaces, uniting the world more and more. Standardized ways of doing things in the scientific and business domains became more and more common. In a phrase, the invention of the printing press paved the way to the establishment of a global civilization (McLuhan 1951, 1962, 1964; McLuhan & McLuhan 1988).

The Alphabet Principle has not only rendered writing less cumbersome, but it has also opened up new representational and communicative possibilities. For example, alphabets can be used to classify things (in alphabetic order) and any letter can be used to represent something in and of itself — think for instance, of the many meanings ascribed today to the letter "X"— X-treme sports, XXX movies; and the list could go and on. The Alphabet Principle has, in effect, become itself a code and, thus, a source of new signification. In a fundamental sense, it parallels in importance the emergence of the decimal system in mathematics, which is based on the concept of place value. The decimal system (from Latin *decem* "ten") uses 10 number signs, also known as *digits* (meaning "fingers" in Latin): 1, 2, 3, 4, 5, 6, 7, 8, 9, and 0. The value of any of these depends on the place it occupies in the numeral. The sign 2, for example, has different values in the numerals 832 and 238, because it is in different places in each: in 832 its value is "two," in 238 its value is "two hundred."

The decimal system has made numerical representation maximally efficient. Not only, but it has made it possible for decimal signifiers to suggest new ideas (signifieds) on their own. For example, the use of superscripts in the representation of exponential numbers has led to the investigation of numerical pattern and to the development of new laws governing numbers. This would have been impossible without such notation. Exponents are shorthand forms for multiplication, making it more efficient for people to multiply repeating digits. Thus, 3^9 is much easier to deal with than is 3 x 3 x 3 x 3 x 3 x 3 x 3 x 3 x 3. Any number, "x" to the exponent "n" can be represented as x^n. At first, this notation was used simply as a way to abbreviate the multiplication of identical digits. But, by investigating the new notation mathematicians came up with new knowledge, such as, for example, that x^0 is always equal to "1". Without going into the simple proof of this here (any basic mathematics textbook will contain it), suffice it to say that it has revealed a property of "0" that would not have been discovered without the notation.

Similarly, alphabets have led to the discovery of features of language and their relation to human cognition that would not have been possible otherwise (Illich & Sanders 1988; Man 2000) — such as the theory of the phoneme. They have also had an effect on how we think and classify information. Just think of how we organize books in libraries, bibliographies, phone lists, and so on and so forth. In effect, the Alphabet Principle has become a guide to cognitive activity in literate peoples today. It has also made abbreviation and acronymy a basic tendency in language evolution. Abbreviated writing seems to have been used by the Greeks as early as the fourth century BC, gradually evolving into a true shorthand code, known as *tachygraphy*. It was the slave Tyro who has been credited with inventing the first true shorthand around 60 BC (after alphabets had become the norm), apparently for recording the speeches of Cicero — a system that was adopted and used widely until the Middle Ages (Cherry 1957: 35). The use of letters to create codes, ciphers, acronyms, and anagrams in the ancient world started to proliferate only after alphabets became the norm (Danesi 2002).

The SCP is particularly evident in how language has developed in cyberspace. Known as "Internetese," the language of cyberspace can be characterized as a condensed Zipfian code with its own type of abbreviations, acronyms, and slang aimed at reducing the time and effort required to relay messages in cyberspace (Crystal 2001): e.g. *b4* = "before," *f2f* = "face-to-face," *gr8* = "great," *h2cus* = "hope to see you soon," *g2g* = "gotta go," and so on. This new code also contains emoticons, as they have come to be called, which are designed to lend compressed iconic commentary to messages: e.g. ;-) = "winking," :-o = "shocked," }:-| = "anger)," and so on. Emoticons are, of course, icons, not alphabet characters, but they can easily be accommodated in an overall framework of analysis, showing the relation between modern-day icons or pictographs (such as emoticons) and alphabetization processes. Suffice it to say here that the characteristics of the Internetese code are now spreading to language generally. Spelling patterns that would have been seen as improper not too long ago are now much more acceptable, thanks to Internetese and to the unconscious force within humans, known as the Principle of Least Effort, that has generated it.

Concluding Remarks

A consideration of the Alphabet Principle from the standpoint of Zipfian analysis suggests that alphabets came about serendipitously as part of a larger psychobiological tendency in human beings to render communication (in the written medium) more compact and efficient. It also reveals that this tendency in no way restricts or diminishes human ingenuity and creativity. Indeed, as discussed in this paper, once alphabet characters came into existence as economic reductions of pictographs they took on a semiotic life of their own—a life that has become rather productive in the language used in cyberspace and in the modern world generally. This kind of Zipfian assessment of change in sign systems probably characterizes how many (if not all) innovations are made in representation and communication generally. The specific agenda for research in Zipfian semiotics consists of a series of general, yet intriguing, questions which, nevertheless, can lead to specific empirical investigations. Among the questions with regard to alphabets the following come to mind, constituting a "sub-agenda" of research in this specific area of semiotics:

- Is there a Zipfian relation among the forms and functions of different alphabets (e.g. Roman vs. Cyrillic) and the number of characters in them?
- Is there a Zipfian relation among font styles (bold, italic, etc.)?
- Is there a Zipfian relation between the abbreviation of words and the abbreviation of sentences (as seen in text messaging and other cyberspace communication styles)?
- Is there any Zipfian relation between the frequency of letters occurring at the start of words and their distribution in other environments?
- Does a Zipfian analysis suggest that the Alphabet Principle is no longer functional, given the various use of acronyms and other such devices in cyberspace environments?

There are many more questions that Zipfian analysis might elicit, and many of these will come about by actually carrying out semiotic research to investigate Zipfian phenomena in sign systems. One of the areas that would certainly fall under the rubric of nonverbal semiotics is the possible application of Zipfian analysis to such domains of communication as gesture, facial expression, tactile communication, haptics, and the like. Is there a relation between gesture form and its utility and frequency? Are gestures that are more abrupt than others higher in frequency (of usage) than are those that are more elaborate? How does one measure "length" or "size" in the nonverbal domain? Do nonverbal forms that are used with verbal ones — such as the gestures that accompany speech — match them in Zipfian terms? The list could go on and on.

In essence, the semiotic agenda, which has been somewhat unforthcoming in theoretical matters over the last few decades, would be revitalized as a consequence of addressing such questions. Zipfian analysis has the capacity to show that humans, in their apparent quest for economy, end up producing new systems that produce new ideas and serendipitous discoveries. What Zipf's Law shows, in effect, is that general conditions exist in sign systems that determine the equilibrium of the systems in terms of their forms and meanings. It is the specific conditions that shift with time and place, not the general semiotic tendencies.

Note

An earlier version of this paper was prepared for the conference titled "Semiotics and Writing" held at the University of Toronto in 2003.

References

ARNHEIM, R. (1969). Visual Thinking, Berkeley: University of California Press.

BILLETER, J.F. (1990). The Chinese Art of Writing, New York: Rizzoli.

BURKE, J. & KINCANNON, E. (1991). Benford's Law and Physical Constants: The Distribution of Initial Digits. *American Journal of Physics* 14: 59-63.

CHERRY, C. (1957). On Human Communication, Cambridge, Mass.: MIT Press.

COULMAS, F. (1989). The Writing Systems of the World, Oxford: Blackwell.

COULMAS, F. (2003). Writing Systems: An Introduction to their Linguistic Analysis, Cambridge: Cambridge University Press.

CRYSTAL, D. (2001). Language and the Internet, Cambridge: Cambridge University Press.

DANESI, M. (2002). The Puzzle Instinct: The Meaning of Puzzles in Human Life, Bloomington: Indiana University Press.

DANIELS, P.T. & BRIGHT, W. (1995) (eds.). The World's Writing Systems, Oxford: Oxford University Press.

DAVIES, W.V. (1988). Egyptian Hieroglyphs, Berkeley: University of California Press.

DRIVER, G.R. (1976). Semitic Writing: From Pictograph to Alphabet, London: Oxford University Press.

DRUCKER, J. (1995). The Alphabetic Labyrinth: The Letters in History and Imagination, London: Thames & Hudson.

ECO, U. (1994). The Search for the Perfect Language, London: Blackwell.

FISCHER, S.R. (2001). A History of Writing, London: Reaktion Books.

GAUR, A. (1984). A History of Writing, New York: Charles Scribner's Sons.

GELB, I.J. (1963). A Study of Writing, Chicago: University of Chicago Press.

HARRIS, R. (1986). The Origin of Writing, London: Duckworth.

HILL, T.P. (1996). The First Digit Phenomenon. *American Scientist* 86: 358-363.

ILLICH, I. & SANDERS, B. (1988). The Alphabetization of the Popular Mind, San Francisco: North Point Press.

KUCERA, H. & FRANCIS, W.N. (1967). Computational Analysis of Present-Day American English, Providence: Brown University Press.

KUHN, T. (1962). The Structure of Scientific Revolutions, Chicago: University of Chicago Press.

LI, W. (1992). Random Texts Exhibit Zipf's-Law-Like Word Frequency Distribution. *IEEE Transactions on Information Theory* 38:1842-1845.

LOGAN, R. K. (1987). The Alphabet Effect, New York: St. Martin's Press.

MAN, J. (2000). Alpha Beta: How 26 Letters Shaped the Western World, New York: John Wiley & Sons.

MANDELBROT, B. (1954). Structure formelle des textes et communication. *Word* 10: 1-27.

MANDELBROT, B. (1983). The Fractal Geometry of Nature, New York: Freeman and Co.
MARTIN, H-J. (1999). The History and Power of Writing, Chicago: University of Chicago Press.
MCLUHAN, M. (1951). The Mechanical Bride: Folklore of Industrial Man, New York: Vanguard.
MCLUHAN, M. (1962). The Gutenberg Galaxy, Toronto: University of Toronto Press.
MCLUHAN, M. (1964). Understanding Media, London: Routledge and Kegan Paul.
MCLUHAN, M. & MCLUHAN, E. (1988). Laws of Media: The New Science, Toronto: University of Toronto Press.
MILLER, G.A. & NEWMAN, E.B. (1958). Tests of a Statistical Explanation of the Rank-Frequency Relation for Words in Written English. *American Journal of Psychology* 71: 209-218.
NOWAK, M.A. (2000). The Basic Reproductive Ratio of a Word, the Maximum Size of a Lexicon. *Journal of Theoretical Biology* 204:179-189.
ONG, W.J. (1977). Interfaces of the Word: Studies in the Evolution of Consciousness and Culture, Ithaca: Cornell University Press.
PERLINE, R. (1996). Zipf's Law, the Central Limit Theorem, and the Random Division of the Unit Interval. *Physical Review* 54:220-223.
POSNER, R. (ed.) (1983). Special Issue "Language, Written Language, Planned Language." *Zeitschrift für Semiotik.*
PSYCHOYOS, D.K. (2005). The Forgotten Art of Isopsephy and the Magic Number KZ. *Semiotica* 154: 157-224.
RAIMI, R.A. (1969). The Peculiar Distribution of First Digits. *Scientific American* 221:109-119.
RIDLEY, D.R. & GONZALES, E.A. (1994). Zipf's Law Extended to Small Samples of Adult Speech. *Perception and Motor Skills* 79:153-154.
ROBINSON, A. (1995). The Story of Writing, London: Thames and Hudson.
ROUSSEAU, R. & ZHANG, Q. (1992). Zipf's Data on the Frequency of Chinese Words Revisited. *Scientometrics* 24: 201-220.
SAMOYAULT, T. (1998). Alphabetical Order: How the alphabet Began, New York: Viking/Penguin.
SCHMANDT-BESSERAT, D. (1978). The Earliest Precursor of Writing. *Scientific American* 238: 50-59.
SCHMANDT-BESSERAT, D. (1992). Before Writing, 2 vols. Austin: University of Texas Press.
SEBEOK, T.A. (2001). Global Semiotics, Bloomington: Indiana University Press.
SEBEOK, T.A. & DANESI, M. (2000). The Forms of Meaning: Modeling Systems Theory and Semiotics, Berlin: Mouton de Gruyter.

WALKER, C.B.F. (1987). Cuneiform, Berkeley: University of California Press.

WATT, W.C. (ed.) (1994). Writing Systems and Cognition: Perspectives from Psychology, Linguistics, and Semiotics. Dordecht: Kluwer Academic Publishers.

WYLLYS, R.E. (1975). Measuring Scientific Prose with Rank-Frequency ('Zipf') Curves: A New Use for an Old Phenomenon. *Proceedings of the American Society for Information Science* 12: 30-31.

ZIPF, G.K. (1935). Psycho-Biology of Languages, Boston: Houghton-Mifflin.

ZIPF, G.K. (1949). Human Behavior and the Principle of Least Effort, Boston: Addison-Wesley.