Available online at www.sciencedirect.com
SCIENCE DIRECT®
BioSystems 73 (2004) 131–139
BioSystems
www.elsevier.com/locate/biosystems

Zipf's law in phonograms and Weibull distribution in ideograms: comparison of English with Japanese

Terutaka Nabeshima [a], Yukio-Pegio Gunji [a,b,*]

[a] *Graduate School of Science & Technology, Kobe University, Kobe, Japan*
[b] *Department of Earth & Planetary Sciences, Faculty of Science, Kobe University, Kobe, Japan*

Received 4 August 2003; received in revised form 13 November 2003; accepted 20 November 2003

Abstract

Frequency distribution of word usage in a word sequence generated by capping is estimated in terms of the number of "hits" in retrieval of web-pages, to evaluate structure of semantics proper not to a particular text but to a language. Especially we compare distribution of English sequences with Japanese ones and obtain that, for English and Japanese phonogram, frequency of word usage against rank follows power-law function with exponent 1 and, for Japanese ideogram, it follows stretched exponential (Weibull distribution) function. We also discuss that such a difference can result from difference of phonogram based- (English) and ideogram-based language (Japanese).

Keywords: Weibull distribution; Phonogram; Ideogram

1. Introduction

Language is one of the most intriguing topics in complex systems. Since Chomsky proposed universal grammar providing a theory of syntactical aspect of language (Chomsky, 1972, 1984), there is a growing literature on evolutionary process of language and language acquisition discussed in the context of evolution (e.g., Pinker, 1979; Bickerton, 1990). Especially in the framework of dynamical systems, there are many mathematical descriptions of language evolution (Batali, 1994; Hashimoto and Ikegami, 1996; Huford et al., 1998; Nowak et al., 2000; Christiansen et al., 2002; Kamarova and Nowak, 2003). These mathematical descriptions are based on syntactical operations originated from Chomsky, and are separated from real semantics of language or cognitive world. Especially for natural language studies, there is a separation between approaches based on generative grammars (syntactical aspect) and approaches based on cognitive linguistics (semantic aspect) (e.g., Lakoff and Núñez, 2000). It is necessary to focus on the relationship between syntactical and semantic aspects because language is a unity as a whole (Kubozono and Oota, 1998).

As a case study on the relationship between syntactical operation and real semantics, we study distribution of words in a generated word sequence, especially focusing on the comparison of Japanese with English. Some languages are based on phonograms, and others are on ideograms, where it is just in the relative sense. It implies that comprehension of English proceeds syntactically first while Japanese semantically first. Because ideograms carry semantic content by themselves, the role of semantics in an ideogram

* Corresponding author.
E-mail address: yukio@kobe-u.ac.jp (Y.-P. Gunji).

0303-2647/$ – see front matter
doi:10.1016/j.biosystems.2003.11.002

sequence is expected to be different from that in a phonogram sequence. We address the question, how the difference between ideogram and phonogram affects semantic structure of a language. We take English as a representative for phonogram and Japanese as an ideogram. English is a grammar-oriented language on the one hand, and Japanese is a context-oriented language on the other hand (Kubozono and Oota, 1998; Ymauchi, 1974; Tokieda, 1941). An English sentence is constituted based on a grammar, and it explicitly designates message content and a speaker as a subject. By contrast, a Japanese sentence is loosely arranged as a nested structure in an implicit context, and even if a subject is omitted then it can be comprehended (Tokieda, 1941). We can assume that an English sentence is operationally arranged by words and then it generates a message (i.e., semantic content), and that a Japanese sentence results from an implicit message and/or context. In English, syntactical operation is prior to semantic context, although in Japanese it is reversed, where it is just relative difference. We assume that difference of a language in a term of phonogram or ideogram can influence the structure of semantics, and then evaluate the assumption.

The question arises how one can estimate semantics, especially quantitatively. For this purpose we address Zipf's law that is frequency of usage of a word against rank in a text (e.g., a novel) (Zipf, 1935, 1949). It was observed that Zipf's plots typically follow a power-law function with exponent close to 1. Because frequency of usage of a word represents the number of possible usages, frequency of usage is expected to be proportional to the number of acceptation of a word in a given text. It is, therefore, reflected in semantic structure of a text. Although Zipf's law becomes a fundamental law in bibliometrics and library sciences (Ikpaahindi, 1985; White and McCain, 1989), it is not argued with respect to the number of acceptation. Recently, Zipf's law is studied for citation of scientific publication (Render, 1998) and web-page access statistics (Cunha et al., 1995). It can be thought that it is reflected in semantic structure in a net-society. The number of acceptations is, therefore, approximately evaluated by the number of "hits" in retrieval in a net-society.

Zipf's law is no longer a stranger to biological and medical fields. It is debated whether oligonucleotide frequency in DNA sequences follows a power-law (Martindale and Konopka, 1996; Israeloff et al., 1996; Bonhoeffer et al., 1996; Voss, 1996), and is also discussed in grading psychiatric patients (Piqueira et al., 1999). More recently, exponent of Zipf's plots is used for cancer classification (Li and Yang, 2002). In an animal's behavioral sequence in learning process, Zipf's plots follow a power-law only in the period of learning completion (Mizukami et al., 1999; Kitabayashi et al., 2001). These observations shows that the structure of Zipf's plots in terms of exponent and/or the range of ranking fitted by a power-law can reflect a structure of dynamics of a system in question.

In order to evaluate semantic structure of English and Japanese, we take word sequences generated by capping. Because capping is based just on either phonogram or ideogram, we can remove the effect of arbitrary context, and can access the universal semantic structure. Secondly, Zipf's plots of word frequency in net-society are obtained by retrieval for web-page, and we evaluate whether it follows either a power-law or stretched exponential law called Weibull distribution (Weibull, 1951; Meeker and Escober, 1998). As a result, we show that there is a clear difference; word sequences in English and Japanese phonogram shows power-law with exponent 1 and those in Japanese ideogram show Weibull distribution. The role of semantics based on our assumption is also discussed with respect to Zipf's plots with power-law or Weibull distribution.

2. Method: word sequences generated by capping and Zipf's plots

A sentence of natural language is generated by a particular rule in referring to context and cognitive world, whether language is based on ideograms or phonograms. We here call such a cognitive world proper to a language real semantics of a language. Zipf's law shows a universal structure of real semantics of language. Looking into a dictionary leads to some acceptations for a word. The number of acceptation corresponds to the number of occasions at which different usages of a word are carried out. That is why frequency of usage of a word can be regarded as frequency of acceptations, and one can estimate structure of natural language semantics. Given a text like a novel, frequency distribution with respect to usage of a word is

obtained, and that reflects structure of semantics of a novel in question and proper to a language.

Our main purpose is to address structure of semantics of natural language, and to detect difference between semantics based on ideograms (Japanese) and that based on phonograms (English). Because phonogram do not carry meaning by itself, syntactical operations plays an essential role in generating a sentence. By contrast, ideogram involves meaning by itself, and then the role of syntactical operation or a generative grammar is expected to be weaker than phonogram. There can be difference between phonogram- and ideogram-based languages with respect to the relationship between syntactical operation and semantics.

In order to detect structure of semantics proper to a language, influences of contexts proper to a text have to be removed as much as possible. For this purpose, we take a word sequence generated by capping, a word game. As for phonogram, capping proceeds by a rule such that last letter of a precedence word coincides with first letter of a subsequent word. In English a letter does not have a unique pronunciation that is determined as a syllable (Kubozono and Oota, 1998). That is why we use two kinds of capping for English, pronunciation capping and letter capping. By pronunciation capping, a word sequence is generated such as mood → day → India. By letter capping, it is generated such as mood → day → year. Japanese consists of ideogram, Chinese characters, and phonogram, hiragana, where hiragana is originated from modification of Chinese characters. Because Japanese phonogram has a unique pronunciation, pronunciation capping exactly coincides with letter capping. As for ideogram, Chinese characters, a word sequence is generated, by using letter capping. A rule of capping is only using nouns and prohibition of the same word in a single game not to remove a loop sequence.

Although capping is a generator of a word sequence based on letter or pronunciation, a sequence generated by capping is influenced by meaning or semantics. As for an ideogram, Chinese character, there are some semantic content carried by a latter. That is why two words connected by letter capping have similar meaning, and a sequence generated by capping is accompanied with an aggregation of meaning that is latent context. Although capping itself has no consistent particular context, a sequence generated by capping can spontaneously poses a context. Such a context can be regarded as semantics proper to a language. Even for phonogram, a word and/or word sequence is generated by a particular rule such that similar syllables are avoided (Ito and Mester, 1986; McCarthy, 1986). Such a rule does not directly influence generating context, but indirectly carry a context proper to a language. As a result, if one takes a word sequence generated by capping, one can remove influences of a particular context proper to a text and can access a universal structure of semantics proper to a language.

We take two kinds of capping, capping by a man and capping by a machine. If a man generates a word sequence by capping, he searches for a subsequent word depending on temporal perpetual context. Therefore, a word sequence is strongly influenced by particular context. Because capping depends on one's vocabulary, the length of a sequence is restricted. As for a machine capping, it is implemented by a particular program. It looks into a subsequent word in the dictionary of PC-KIMMO developed by the Summer Institute of Linguistics (http://www.sil.org/). Searching items are restricted in a set of appellatives. By using a machine capping, one can obtain a long word sequence, however it is restricted under a rule of capping.

For a word sequence generated by capping, frequency of a word usage in a net-society is estimated. By using a retrieval tool (http://www.google.co.jp/), the number of web-pages involving a word is calculated as frequency of usage of a word in question. After that, Zipf's plots (i.e., frequency of word usage against rank) are obtained. The top measurement is ranked the first, or rank $= 1$, and second best measurement has rank $= 2$, etc., and then one obtains a monotonously decreasing function as a distribution. We examine whether Zipf's plots follow a power-law or follow a Weibull distribution.

3. Power-law or Weibull distribution in frequency of word usage

We estimate whether Zipf's plots of a word sequence follow a power-law or not. In addition, we estimate whether they follow a stretched exponential law, Weibull distribution. First we show Zipf's plots of word sequences generated by capping in English. As mentioned before, we obtain two kinds of word

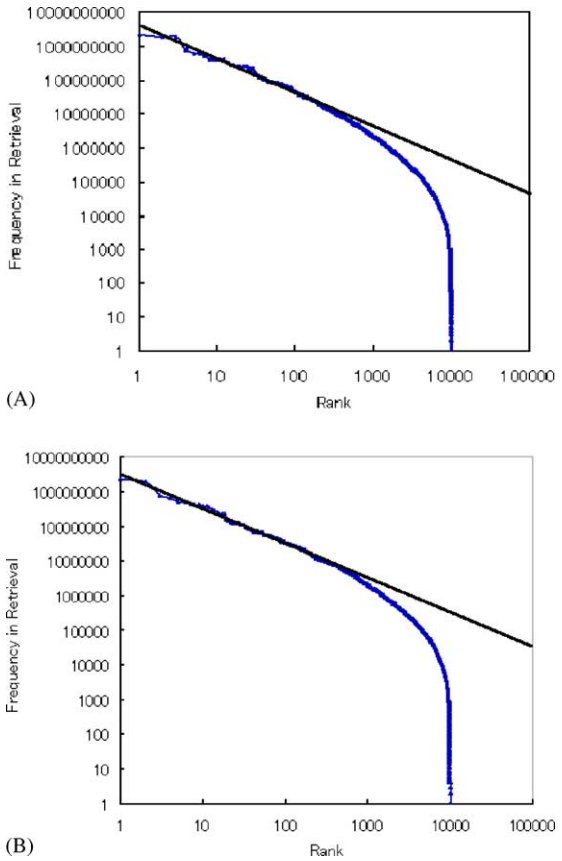

Fig. 1. Frequency of word usage vs. rank (Zipf's plot) in an English sequence generated by machine capping, where it is plotted in log-log scale. Frequency is evaluated by the number of "hits" web-pages in retrieval for a word. (A) Word sequence generated by pronunciation capping. (B) Word sequence generated by letter capping. Line represents the slope of −1.

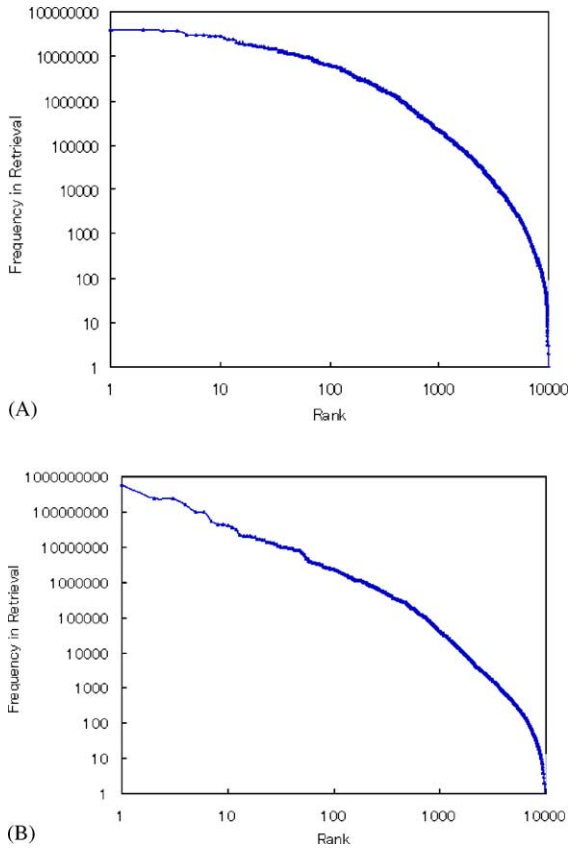

Fig. 2. Frequency of word usage against rank in log-log plot for a Japanese sequence generated by machine. (A) Zipf's plots for Chinese character sequence. (B) Zipf's plot for Japanese phonogram, hiragana.

sequences; generated by letter capping and by pronunciation capping. Fig. 1 shows Zipf's plots of a word sequence generated by machine capping. A word sequence is generated by capping till the sum of words becomes 10,000. Frequency of a word usage in a net-society is plotted against the rank of a word, in log-log scale. It is clear for both sequences generated by letter capping and pronunciation capping that Zipf's plots can be fitted by a power-law function through the range of high ranking words, over three decades. The exponent is close to 1, and that coincides with Zipf's law (Zipf, 1935, 1949).

Zipf's plots for Japanese ideogram (Chinese character), however, do not show power-law, whether a word sequence is generated by machine capping or by man capping. Fig. 2 shows frequency of word usage against rank in log-log plot, where a word sequence is generated by machine. Zipf's plots for Chinese character sequence (Fig. 2A) cannot be fitted by a power-law function. By contrast, those for Japanese phonogram (hiragana) can be fitted by a power-law function in the range of high ranking words (Fig. 2B).

Instead of a power-law function, it is estimated whether Zipf's plots can be fitted by Weibull distribution. It is well known as the probability of failure, y, at time, x that is expressed as

$$y = \alpha\beta x^{\beta-1} \exp(-\alpha x^{\beta})$$

with parameters, α and β (Weibull, 1951; Meeker and Escober, 1998). We replace variables time and

probability of failure by rank and normalized frequency of word usage, respectively. Given a probability distribution function as Weibull distribution, corresponding cumulative probability distributive function is expressed as $F(x) = 1 - \exp(\alpha x^{\beta})$, and then we obtain that $\log(\log(1 - F(x))) = \beta \log(x) + \log(\alpha)$. Then a straight line with slope β is obtained in log(log)-log plot.

Fig. 3 shows normalized cumulative frequency of a word usage against rank in log(log)-log plot, where a word sequence consists of Japanese, Chinese character (Fig. 3A) and hiragana (Fig. 3B), and both of them are generated by machine capping. In the range of large rank a straight line can be observed. It implies that Zipf's plots can be fitted by Weibull distribution function in the range of low ranking usages. The slope of regression line is 0.2457 for hiragana, and is 0.2531 for Chinese characters. We also estimate whether English sequence of capping follows Weibull distribution or not. Fig. 3C and D shows the plots for English, where data of word sequences are the same as ones used in Fig. 1. They do not follow Weibull distribution.

Japanese word sequences are generated also by man's capping. Fig. 4A and B shows Zipf's plots in log-log plot, and Fig. 4C and D shows normalized cumulative frequency of a word usage against rank in log(log)-log plot. Especially for Chinese character

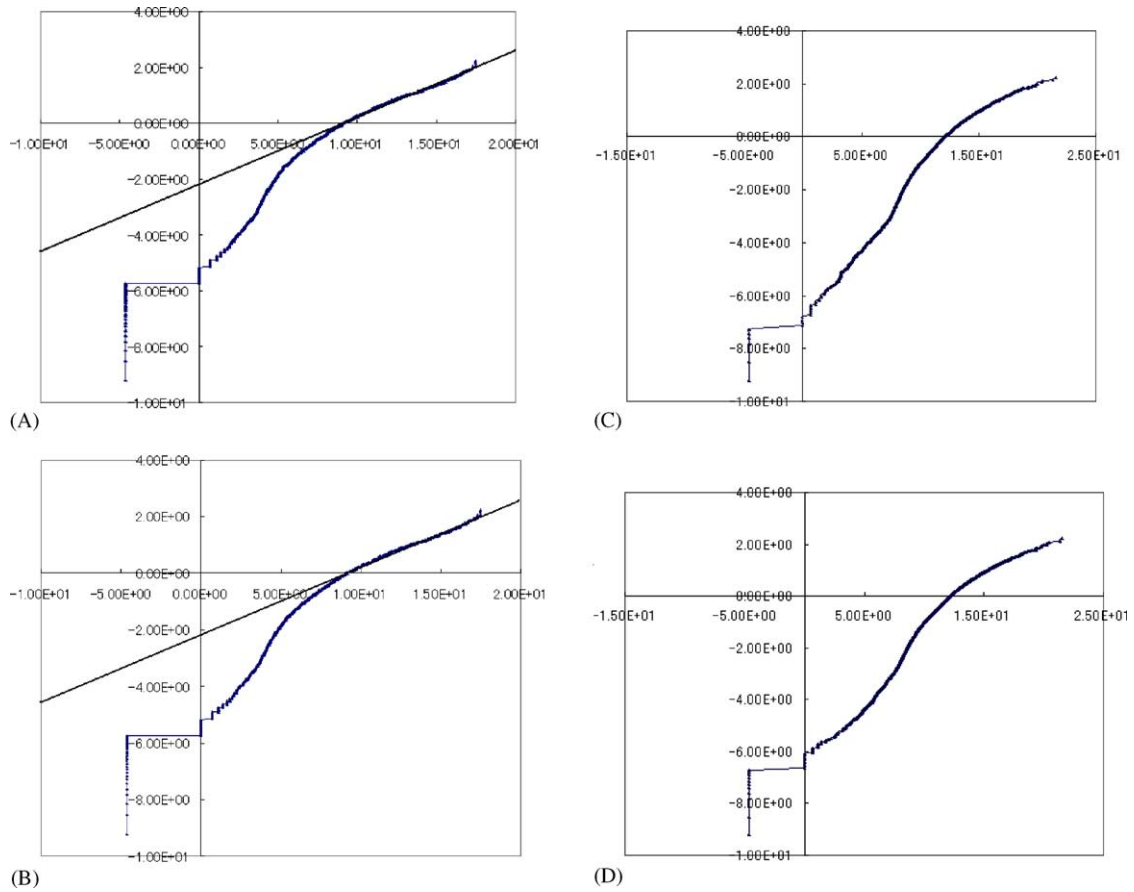

Fig. 3. Normalized cumulative frequency of a word usage against rank in log(log)-log plot. If it can be fitted by a line, it shows that Zipf's plot can be fitted by Weibull distribution function. A word sequence consists of Japanese, Chinese character (A) and hiragana (B). By contrast a word sequence consists of English, and is made by letter capping (C) and pronunciation capping (D). All of word sequences are generated by machine capping.

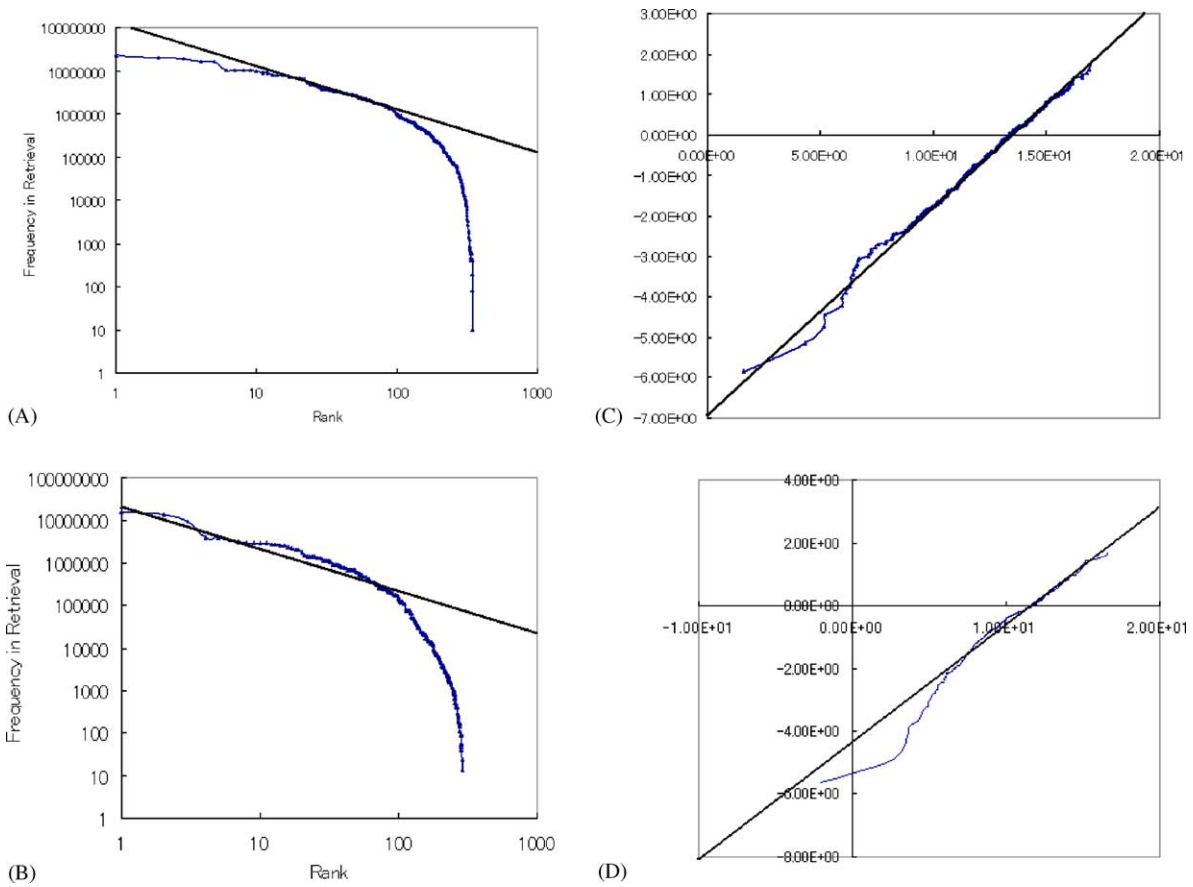

Fig. 4. Zipf's plots for Japanese word sequences generated also by man's capping. (A) Zipf's plots for Chinese characters (ideogram) in log-log plot. (B) Zipf's plots for hiragana (phonogram) in log-log plot. (C) Normalized cumulative frequency of a word usage against rank in log(log)-log plot, corresponding to (A) and (D). Normalized cumulative frequency of a word usage against rank in log(log)-log plot, corresponding to (B).

sequence, plots can be fitted by Weibull distribution function through whole range of ranking (Fig. 4C). It may be relevant for the property of ideogram, and for what the role of semantics is prior to syntactical operation.

There is clear distinction between English and Japanese ideogram with respect to Zipf's plots (frequency of word usage against rank). An English word sequence generated by capping follows a power-law. By contrast, a Japanese ideogram sequence does not follow a power-law and Zipf's plot can be fitted by Weibull distribution function rather than power-law function. Such a difference might result from a difference between a language based on phonogram and that based on ideogram, and that is based on the fact; Zipf's plots for Japanese phonogram follows a power-law with exponent 1. Especially Japanese phonogram also poses a character of Weibull distribution for large ranking. It might be explained that hiragana is originated from modifications of Chinese character (Figs. 2B and 3B).

A formal language is described as syntax (generative grammar) and as semantics (a model of syntax). Syntactical description is isomorphic to semantic description, and then one does not have to describe a language by both of them. It is satisfactory to define a language only by one of either syntax or semantics (Goldblatt, 1991). By contrast, natural language consists both of generative grammar and cognitive world (Kubozono and Oota, 1998). Meaning carried by a

word is determined influenced by cognitive world. In this sense cognitive world corresponds to semantics in formal language. The essential difference between formal and natural language is whether semantics is definite or indefinite. Although relatively generative grammar can be explicitly described, semantics of natural language cannot be explicitly described because it is indefinite. Therefore, explicit expression for semantics of natural language is destined to be just an approximation. That is why generative grammar (syntax) does not correspond to cognitive world (semantics) in an exact sense. Syntax and semantics have different roles in natural language. It is necessary to evaluate the relationship between generative grammar and semantics.

Different language might have different relationship between generative grammars and semantics. If a language is based on phonogram, a letter itself cannot carry meaning or semantic content. In that language, the role of semantics is relatively small in generating a sentence, and generative grammar is dominant. By contrast, if a language is based on ideogram, the role of generative grammar is relatively small in generating a sentence, and it looks as if there existed semantics like a potential function. It leads to a hypothesis that generative grammar is dominant in phonogram-based languages and that semantics like a potential is dominant in ideogram-based languages.

If a generative grammar is dominant in a language, a word generates another word by following a generative grammar, and it leads to a long sentence. Perpetual genesis of words proceeds like a catalytic network with a cascade fashion (Aizawa, 1998), and reaches a steady state in a long enough sentence. In the situation, the relationship between rank, x, and frequency of word usage, y, is expected to be expressed as

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \alpha\left(\frac{y}{x}\right).$$

Because lower ranking word generates higher ranking words dependent on their frequencies, and that is balanced in a steady state. It leads to a power-law, $y = x^{\alpha}$.

If semantic is dominant in a language, one can approximately assume a particular function by which information content originally carried by a word is determined. That information is assumed to be intentional one by which a various meaning can be derived (Tokieda, 1941). Imagine the following situation. When one walks in a mountain, he takes a branch of tree and says that this is a good walking staff. In this case a word "walking staff" indicates a branch. In other words, a walking staff inspires a material, branch, that is not a walking staff (Tokieda, 1941). The information mentioned here is regarded as a force penetrating different metaphoric level (e.g., walking staff $\rightarrow$ branch). If it is assumed that the deviation of information is proportional to itself, it is expressed as

$$\frac{\mathrm{d}f(u)}{\mathrm{d}u} = f(u),$$

where u is a particular variable corresponding a word, and is arranged by similar information. Because the information, $f(u)$, represents a driving force penetrating different metaphoric level, word acceptation of a word with rank, x, is expected to be dependent on the deviation of $f(u)$. If $f(u)$ and $f(u + \Delta u)$ are the same, they inspire the same metaphorical level and then acceptation of a word, u, is very low. The difference between $f(u)$ and $f(u + \Delta u)$ per Δu can contribute acceptation of a word, u. It is also assumed that the relationship between rank, x, and u is expressed as,

$$u = -\alpha x^{\beta}.$$

From the assumption, the number of acceptation of a word, y, with rank x is expressed as

$$y = -\left(\frac{\mathrm{d}f(u)}{\mathrm{d}u}\right)\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right) = -\left(\frac{\mathrm{d}u}{\mathrm{d}x}\right)\exp(u)$$
$$= \alpha\beta x^{\beta-1}\exp(-\alpha x^{\beta}).$$

It is Weibull distribution function. Therefore, if a word originally carries information inspiring metaphoric usage, as discussed by Tokieda (1941), frequency of usage of a word is expected to follow a stretched exponential distribution (at least exponential) rather than power-law.

4. Conclusion

To study the relationship between syntactical operation and semantics, we analyze frequency of a word usage in net-society (Zipf's plot). Because Zipf's plots represent the distribution of the number of possible usages of words, they show distribution of word acceptations and/or structure of semantics. Especially

we focus on, how syntactical operation influences the structure of semantics. For this purpose, we compare a language based on phonogram with one based on ideogram, and compare English with Japanese.

A word sequence generated by capping is analyzed to remove the influence of context proper to a particular text, and two kinds of capping, man's capping and machine capping are conducted. As for machine capping, a word sequence is generated with the length of 10,000 words by a particular program referring to a dictionary. Frequency of word usage in a net-society is estimated by retrieval of web-pages, and it is examined whether distribution follows power-law or not. If distribution does not follow a power-law, it is estimated whether it follows Weibull distribution function or not. As a result, we obtain that Zipf's plots of English sequences and Japanse phonogram sequences follow power-law with exponent close to 1 (i.e., Zipf's law), and that Zipf's plots of Japanese ideogram sequence follow Weibull distribution function. As for man's capping, such a distinction, power-law in phonogram and Weibull distribution in ideogram is clearer than the case of word sequences by machine capping.

We also suggest that a language which syntactical operation is prior to semantics show Zipf's plots following power-law, and that a language which semantics is prior to syntactical operation show Zipf's plot following stretched exponential law. A language consisting only of phonogram can be based on syntactical operation rather than based on semantics, compared with a language consisting of ideograms. It is consistent with our analytical results that English (phonogram) sequences show Zipf's plot with power-law and Japanese (ideogram) sequences show Zipf's plot with Weibull distribution function.

References

Aizawa, Y., 1998. Unbroken wholeness in nonlinear processes. Int. J. Comput. Anticipatory Syst. 2, 235–249.

Batali, J., 1994. Innate biases and critical periods. In: Brooks, R., Maes, P. (Eds.), Artificial Life, vol. IV. MIT Press, Cambridge, MA, pp. 160–171.

Bickerton, D., 1990. Language and Species. University of Chicago Press, Chicago.

Bonhoeffer, S., Herz, A.V.M., Boerlijst, M.C., Nee, S., Nowak, M.A., May, R.M., 1996. Explaining 'linguistic features' of noncoding DNA. Science 271, 14–15.

Chomsky, N., 1972. Language and Mind. Harcourt Brace Jovanovich, New York.

Chomsky, N., 1984. Principles and parameters in syntactic theory. In: Horstein, N., Lightfoot, D. (Eds.), Explanation in Linguistics. Longman, London, pp. 123–146.

Christiansen, M.H., Dale, R.A.C., Ellefson, M.R., Conway, C.M., 2002. The role of sequential learning in language evolution: communicational and experimental studies. In: Cangelosi, A., Parisi, D. (Eds.), Simulating the Evolution of Language. Springer, London, pp. 165–187.

Cunha, C.R., Bestavros, A., Crovella, M.E., 1995. Characteristics of WWW client-based traces. Technical Report 95-010, Computer Science Department, Boston University.

Goldblatt, R., 1991. Topoi. The Categorical Analysis of Logic. North-Holland, Amsterdam, 1991.

Hashimoto, T., Ikegami, T., 1996. Emergence of netgrammar in communicating agents. Biosystems 38, 1–14.

Huford, J.R., Studdert-Kennedy, M., Knight, C., 1998. Approaches to the Evolution of Langauge. Cambridge University Press, Cambridge, MA.

Ikpaahindi, L., 1985. An overview of bibliometrics: its measurements, laws and their applications. Libri 35, 153–177.

Israeloff, N.E., Kagalenko, M., Chan, K., 1996. Can Zipf distinguish languages from noise in nocoding DNA? Phys. Rev. Lett. 76 (11), 1976.

Ito, J., Mester, A., 1986. The phonology of voicing in Japanese: theoretical consequences for morphological accessibility. Linguistic Inquiry 17, 49–73.

Kamarova, N.L., Nowak, M.A., 2003. Language dynamics in finite populations. J. Theor. Biol. 221, 445–457.

Kitabayashi, N., Kusunoki, Y., Gunji, Y-P., 2001. The logical jump in shell changing in hermit crab and tool experiment in the ants. In: Yamakawa, T., Matsumoto, G. (Eds.), What Should be Computed to Understand and Make Brain Functions? World Scientific, Singapore, pp. 183–205.

Kubozono, H., Oota, S., 1998. Phonological Structure and Accents. Kenkyusha Publishing Co., Tokyo (in Japanese).

Lakoff, G., Núñez, R.E., 2000. Where Mathematics Comes From? In: How the Embodied Mind Brings Mathematics into Being. Basic Books, New York.

Li, W., Yang, Y., 2002. Zipf's law in importance for cancer classification using microarray data. J. Theor. Biol. 219, 539–551.

Martindale, C., Konopka, A.K., 1996. Oligonucleotide frequencies in DNA follow a Yule distribution. Comput. Chem. 20, 35–38.

McCarthy, J.J., 1986. OCP effects: gemination and antigemination. Linguistic Inquiry 17, 207–263.

Meeker, W.Q., Escober, L.A., 1998. Statistical Methods for Reliability Data. John Wiley & Sons, New York.

Mizukami, E., Migita, M., Gunji, Y.-P., 1999. Self-similar pattern in conceptualization in goldfish. Biosystems 54, 91–104.

Nowak, M.A., Plotkin, J.B., Jansen, V.A., 2000. The evolution of syntactic communication. Nature 404, 495–498.

Pinker, S., 1979. Formal models of language learning. Cognition 7, 217–283.

Piqueira, J.R., Monteiro, L.H., de Magalhaes, T.M., Ramos, R.T., Sassi, R.B., Cruz, E.G., 1999. Zipf's law organizes a psychiatric ward. J. Theor. Biol. 198, 439–443.

Render, S., 1998. How popular is your paper? An empirical study of the citation distribution. Eyr. Phys. J. B4, 131–134.
Tokieda, M., 1941. Principles of Japanese Linguistsis. Iwanami Publishing Co., Tokyo (in Japanese).
Voss, R.F., 1996. Comment on “linguistic features of noncoding DNA sequences”. Phys. Rev. Lett. 76 (11), 1978.
Weibull, W., 1951. A statistical distribution function of wide applicability. J. Appl. Mech. 18, 293–297.
White, H.D., McCain, K.W., 1989. Bibliometrics. Ann. Rev. Inform. Sci. Technol. 24, 119–186.
Ymauchi, T., 1974. Logos and Lemma. Iwanami Publishing Co., Tokyo (in Japanese).
Zipf, G.F., 1935. Psycho-Biology of Languages. Houghton-Mifflin, Boston, MA.
Zipf, G.F., 1949. Human Behavior and the Principle of Least Effect. Addison-Wesley, New York.