

日本語の文章における句読点のゆらぎの解析に関する研究

A Study on Fluctuation Analysis of Punctuation Marks in Japanese Writings

亀井 栄治

EIJI KAMEI

名城大学 都市情報学部

Faculty of Urban Science, Meijo University

1. 研究の目的

日本語の文章は、句読点によって区切られている。これは「読ませるための工夫」であり、英語などの他の言語の文章には見られない、日本語の文章に独特の特徴であるといえる。いうまでもなく、句読点の打ち方は、書き手が決定することのできるものであり、これは日本語に許された自由度であるといえる。このことは、日本語の文章が多分の冗長性を含んでいることを示唆し、句読点の位置によって、文章の意味や内容が変わることも少なくない。したがって、日本語の文章において、句読点のもつ役割は極めて大きいといえる。さらに、句読点の打ち方が、その文章の雰囲気的特点を決定し、書き手の作風が成り立っていることも少なくない。

本研究では、日本語の文章において、句読点の配置によって生じる文章の区切りの長さに着目し、そのゆらぎの解析を行うことにより、文章のもつ特徴を定量的に把握する手法を提案する。

2. ゆらぎ・ゆらぎ解析

ここでは、ゆらぎの概念と、ゆらぎを解析するための基本的流れについて説明する。ゆらぎ解析において対象となる解析データは、数値により記述される変化のデータであり、折れ線グラフなどの「波形」として、視覚的に表現できるものである。

まず、解析データをフーリエ変換 (Fourier Transform) により処理し、各周波数に対するパワースペクトル密度の分布の相関を測定する。その詳細と具体的な手順は 3. で述べるが、得られた結果は、通常、図 1 のようなグラフとして表現される。なお、図 1 のグラフの両軸は対数軸である。このグラフ上でプロットされたパワースペクトル密度の分布 (以下、スペクトル分布という) の傾きは、ゆらぎの程度を表している。すなわち、この傾きは、含まれている周波数の低い波 (低周波) と周波数の高い波 (高周波) との「バランス」を表現しており、このことはいい換えれば、解析データの高周波 (部分的な特徴) と低周波 (全体にかかわる特徴) をともに把握していることになる。ゆらぎを測定するとは、このスペクトル分布の傾きを求めることである。

ここで、いろいろに解析においてよく現れるゆらぎの型について説明する。それらは大別して 3 つに分かれ、それぞれスペクトル分布の傾きが 0, -1, -2 のものである。この様子を図 2 に示す。もちろん、実際の解析結果では、スペクトル分布は散らばりながら、大まかな様子として、これらに近く分布することになる。傾きがこれらの中間にくるような解析結果が得られることは数多いが、ここに紹介する 3 つの型は、特に現れることが多い。スペクトル分布の傾きが 0 であるものは、「白色ゆらぎ (ホワイトノイズ)」とよばれ、どの周波数の波も同じ程度に含まれ、結果として変化が唐突で、意外性に富むデータである。このデータの変化は乱雑であり、したがって、次の様子を予測することは不可能である。スペクトル分布

の傾きが -2 であるものは、「 $1/f^2$ ゆらぎ」とよばれ、低周波の度合いが比較的高く、結果として全体的な変化の様子がはっきりして、これにより変化が単調で、期待性（予測性）に富むデータである。スペクトル分布の傾きが -1 であるものは、「 $1/f$ ゆらぎ」とよばれ、低周波と高周波が適度に混ざって含まれた結果、生じるゆらぎである。この場合は、変化に意外性と期待性（予測性）をあわせもつデータである。 $1/f$ ゆらぎは、快適感や心地よさといった、人のもつ感覚と相関が高いことが、最近の多くの研究で明らかにされてきている。

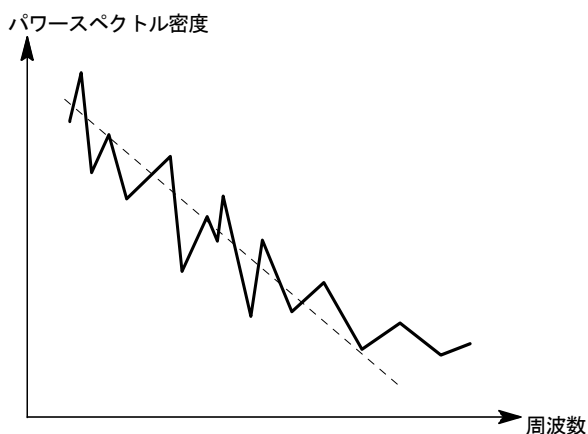


図 1 ゆらぎ

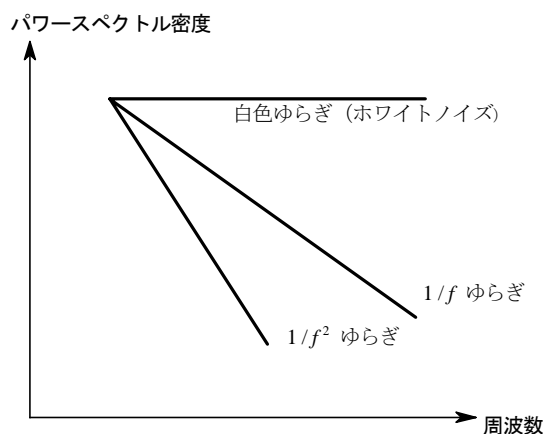


図 2 ゆらぎの型

3. 解析の方法と手順

本研究における句読点のゆらぎ解析の方法と手順は、以下の通りである。

- 1) 解析対象となる日本語の文章をひらがなで記述し、句読点から句読点までを 1 つの区切りと考え、各区切りに含まれる文字数を順に数え、ゆらぎ解析の対象となる解析データを作成する。この際、促音・拗音・長音は 1 文字として数える。ひらがな文を用いる理由は、句読点が「読ませるための工夫」であることから、実際にその文を読む際の区切りの長さを把握する必要があり、漢字かな交じり文のままでは、それができないことによる。その一例を以下に示す。

(例) 今日は雨が降り、気温が低かった。天気予報では、明日は天候が回復し、気温が上がるそうだ。

↓

きょうはあめがふり、きおんがひくかった。てんきよほうでは、あしたはてんこうがかいふくし、きおんがあがるそうだ。

↓

解析データ 8 9 8 14 10

- 2) 解析データに対し、離散的フーリエ変換 (DFT : Discrete Fourier Transform) により、各周波数に対するパワースペクトル密度を計算する。
- 3) 両対数グラフの横軸に周波数を、縦軸にパワースペクトル密度をそれぞれとり、計算結果をプロットし、それらの分布の相関を最小 2 乗法を用いて直線に回帰させることにより表現する。2. で示したように、回帰直線の傾きがゆらぎの程度を示し、本研究では「ゆらぎの値」とよぶ。なお、本研究の解析では、回帰が妥当なものかを判断するために、適合度の検定を χ^2 検定により、危険率 5 %で行い、回帰の妥当

性を確保している。

また、本研究では、2) と 3) の一連の処理を実現する解析プログラムを、C 言語により作成して用いている。

4. ケーススタディ

本稿では、ケーススタディとして、同じ内容を記述した以下の 2 つの文章に対して、ゆらぎ解析を行う。サンプル文 A とサンプル文 B とでは、B の方が句読点の量が多い。それぞれに対し、3. の 1) の処理で得られた解析データと、得られたゆらぎの値を示す。

サンプル文 A

ペットを飼う人が増えれば、ペット用品が多種出回り、その業界のニーズは高まる。近年ペット市場は成長を続け、ついに一兆円市場にまでなった。犬を飼育している世帯数は、一千万件を超えるまでの多さになり、犬猫合計の世帯数ではおよそ二倍となる。今後予想されうることの一つには、『ペット可』の賃貸住宅の建設ラッシュがある。もう一つの予想されうることとして、ペット市場の拡大により、ペットロスがあげられる。ペットが死んだときの精神的ストレスにより、ペットロスの社会問題が起こる。動物愛護国のようなが、事実は異なる。ペットブームの陰に隠れているが、犬は年十七万匹、猫は年二十八万匹もの数の生命が、炭酸ガスにより保健所で殺処分されている。処分で使う炭酸ガスは二酸化炭素ガスのことで、一般的に安楽死だと思われているが、二酸化炭素ガスにより窒息死させられているのが現実だ。また人に飼われていた動物ほどガス室を嫌い、入り口で必死に踏ん張るのだ。動物のガス室での苦しみをを知る職員は、毎日の施設での仕事に胸を痛めている。ペットブームによってペット市場が急速に拡大する裏で、殺処分にかかる莫大な費用に税金が使われているということが、現代社会が抱える矛盾を浮き彫りにしているという結果をもたらす。

解析データ 13 14 16 20 18 16 21 25 18 24 18 14 11 23 17 14 8 17 13 24 24 27 22 31 25
15 23 23 31 36 36

ゆらぎの値 -0.82

サンプル文 B

ペットを飼う人が増えれば、ペット用品が多種出回り、その業界のニーズは高まる。近年、ペット市場は成長を続け、ついに、一兆円市場にまでなった。犬を飼育している世帯数は、一千万件を超えるまでの多さになり、犬猫合計の世帯数では、およそ二倍となる。今後予想されうることの一つには、『ペット可』の賃貸住宅の建設ラッシュがある。もう一つの予想されうることとして、ペット市場の拡大により、ペットロスがあげられる。ペットが死んだときの精神的ストレスにより、ペットロスの社会問題が起こる。動物愛護国のようなが、事実は異なる。ペットブームの陰に隠れているが、犬は年十七万匹、猫は年二十八万匹もの数の生命が、炭酸ガスにより、保健所で殺処分されている。処分で使う炭酸ガスは、二酸化炭素ガスのことで、一般的に、安楽死だと思われているが、二酸化炭素ガスにより、窒息死させられているのが現実だ。また人に飼われていた動物ほどガス室を嫌い、入り口で必死に踏ん張るのだ。動物のガス室での苦しみをを知る職員は、毎日の施設での仕事に胸を痛めている。ペットブームによって、ペット市場が急速に拡大する裏で、殺処分にかかる莫大な費用に、税金が使われているということが、現代社会が抱える矛盾を、浮き彫りにしているという結果をもたらす。

解析データ 13 14 16 4 16 3 15 16 21 16 9 18 24 18 14 11 23 17 14 8 17 13 24 9 15 14
13 7 15 12 19 25 15 23 23 10 21 18 18 16 20

ゆらぎの値 -0.05

解析結果を見ると、サンプル文 A の文章は、かなり $1/f$ ゆらぎに近く、サンプル文 B の文章は、ほぼ白色ゆらぎであることがわかる。どちらの文章が「読みやすい」かは、個々人のもつ感覚によって異なるが、本研究では、サンプル文 B のように、句読点により、細かく区切られている文章では、ゆらぎの値が 0 に近いという結果が得られた。

5. 結論・今後の課題

本研究では、日本語の文章において、句読点の配置によって生じる文章の区切りの長さに注目し、そのゆらぎの解析を行うことにより、文章のもつ特徴を定量的に把握する手法を提案した。本稿において示したケーススタディはその一例であるが、研究の結果では、内容が同じでありながら、句読点の打ち方によって生じる、文章のもつ書き手の「性格」的な面を、一部とはいえども定量的に把握することができることがわかった。

邪推であるが、「5 7 5」や「5 7 5 7 7」で生じる俳句や短歌のリズムは、 $1/f$ ゆらぎに当てはまるのではないかと思う。全体的な基調をもちつつ、少しだけ適度に変化するリズムこそ、 $1/f$ ゆらぎのもつ特徴ではないかと考えるからである。残念ながら、データ数が少ない故に、それらのゆらぎの値を求めることは不可能であるが、それぞれに、心地よい（日本人にとって）独特のリズムがあることは、「何かしら」否めない。本研究の課題として、多量の文章に対して、文章の読み手の印象と、ゆらぎの値との相関関係を心理的に調査することが挙げられる。それによって、本研究で得られた成果は、より精度を増していくものとする。