

# 日本語と英語の画数による比較研究

——文字の画数と情報量——

## A Comparison between Japanese and English in terms of Strokes

横原 恭士

### はじめに

人間は環境からの情報を五感によって感知している。他の人間や情報機器などからの情報は音声、文字、パターンの形で聴覚と視覚によって取入れている。人の脳による情報処理の結果の情報を、人は言葉や文字で表している。また人は情報機械による情報を数字や文字や図表や音声で表している。人や情報機械からの文字や図表により表現できる情報を人間は主に視覚から取り入れることができる。言葉により一次元時系列的に表される情報を文字によって視覚から得ることができる。文字は一次元的時系列的に情報を表すが、文章として完成されたものは二次元的な情報伝達の形態をとる。

視覚は本来二次元的に視野内の情報を捉えるセンサの役割を持つものである。情報量の増加は映像化時代の到来に見られるように、最も情報量を多く取り入れることができる視覚の活躍の場を増やすことになる。

言語には英語のように耳と目から入る情報の形態が共に一次元的であるものと、中国語の漢字のように耳からは一次元的だが目からは二次元的に情報が入るものがある。平仮名、片仮名と漢字を用いる日本語は両言語の混合語の性格を持つと言える。

本稿では一次元的な文字だけを使用する英語と、仮名という一次元的な文字と漢字という二次元的な文字の両方を併用する日本語を、文字数や画数や情報量の点から比較し考察を加える。

### 1 基本的な漢字と英単語

人間の脳の情報処理に関わる情報伝達では「実体」(=ENTITY)に関する情報が重要である。人間間の言語や文字による情報伝達では、主語や目的語が明確でないと正しく

情報は伝わりにくい。主語や目的語の指す「実体」が何であるかを正確に伝えることが重要である。また、「実体」の性質、タイプなどの属性と「実体」と「実体」の関係を正しく伝えることも重要である。このように、言語や文字による情報伝達では、「実体」そのものや「実体」の属性、「実体」と「実体」の関係を表す言葉や文字は重要である。

文字による情報伝達を考えてみると、英語では「実体」や「実体」を表す単語は全てアルファベットを一次元に並べたものである。一方日本語では、これらの単語は殆ど漢字で表わされている。

日本語と英語の比較を、「実体」そのもの、「実体」の属性および「実体」と「実体」の関係を表す単語について行う。

### 1. 1 「実体」に関する漢字の画数と英単語の字数の比較

日本語の起源は漢字であり、漢字は象形文字を起源としている。象形文字は動植物を含む自然や人体の一部や生活に必要な物や状態を表している。

言語による情報伝達が「実体」についての記述であるということと、漢字の起源であり基本である象形文字に「実体」、「実体」と「実体」との関係や「実体」の属性を表すものが多いということから、日本語の中の「実体」に関する象形文字に分類される漢字とそれに対応する英単語をとりあげ比較する。

象形文字より後の時代に考えだされた漢字や、日本で考案された平仮名や片仮名は、「実体」についての説明や補足をするためのものである。日本語の文章ではいろいろな情報を数千の漢字と平仮名、片仮名および場合によっては外来語をそのまま使用することによって表現する。しかし、日本語の基本は「実体」に関する象形文字である。「実体」に関する象形文字である漢字に日本語の本質の一つがあるのである。

一方、英語はAからZの26文字の組合せで単語をつくり、単語間と文章間に空白を配置して文章を構成している。単語で表すものが「実体」(=ENTITY)であれどのような品詞であれ26文字と空白の組合せでそれを表現しているのである。英語は少ない基本文字の組合せで情報を表現することを追求した結果の簡潔な言葉である。英語の場合も、「実体」についての情報を伝えることが、情報伝達上の大切な使命であり、日本語の場合と同様に、「実体」に関する単語は重要な役割を担っている。

情報伝達において、情報表現の対象となっている「実体」に関する文字は日本語では漢字であり、英語ではアルファベットを組み合わせた単語である。文字間の区別を漢字では縦、横、斜めの線を画として組み合わせることによって行っているが、英語では26個の文字と空白を一次的に配列することによって行っている。

表音文字同士の比較、例えば日本語の仮名だけの文章と英語の文章間の比較では各々の基本文字数が平均情報量に関わってくる。この時の実際の平均情報量は、清濁空白の72文

字を基本文字とした時の仮名日本語が1.5ビット、英語が1.0ビット程度である(横原,1991)。このことには、表音文字が一次元に文字を並べて情報を表すことと基本文字数および各言語の文字間のつながりなどが関係している。

漢字では線素の2次元の組合せにより文字の区別を行っている。一次元的に情報を表現する文字であるアルファベットと二次元的に情報を表す文字である漢字の比較を、基本的な漢字の画数と対応する英単語の文字数について行った。漢字の中で基本的なものとして、小学校で出てくる象形文字に分類されるものを選んだ。但し、対応する英単語で熟語になるものや「実体」(=ENTITY) と考えにくいものは除外した。

## 1. 2 比較結果

表1は比較結果である。漢字の平均画数と英単語の平均字数を比べると、平均画数5.4に対し平均字数が4.6である。日本語では漢字の前後に空白はなく英語では単語の前後に空白が必ずあるので、英単語の字数に空白を加えた値を比べると、画数の方が5.4、字数の方が5.6になりかなり近い値となる。この結果から、基本的な「実体」を表す文字であり象形文字に分類される漢字と、対応する英単語を空白の字数を考慮して比較すると、漢字の画数と英単語の文字数に1対1に近い対応がみられることがわかる。

表1 漢字(象形文字)画数と英単語文字数

漢 字			英 語		
象形文字数	画数合計	平均画数	単語数	字数	平均字数
42	227	5.4	42	192	4.6

## 2. 日本語と英語の文章比較

前節では「実体」に関係し象形文字に分類される漢字とそれに対応する英単語について、漢字の画数と英単語の文字数の比較を行なった。

情報伝達では「実体」に関する単語とともにそれらを補足する単語も情報を正確に伝える上から重要である。日本語ではこれらの単語に使用される漢字は象形文字以外に分類されるものの数が多い。平仮名や片仮名も日本語では覚えるべき漢字数の負担を軽減し(横原,1991)、表現に変化をつけるなど日本語において重要な役割を果たしている。英語ではアルファベットと空白の組合せで全ての情報を表現するのに対し、日本語では「実体」「属性」「タイプ」などを表す漢字と平仮名片仮名との組合せで情報を表現している。日本語と英語を情報伝達の観点から定量的に評価するためには文章全体の比較を行っておく必要がある。そのため同じ内容を表し、表現力にあまり差がなく冗長性が同じ程度と思われる

英語と日本語の対訳文について、文章全体の字数、画数および情報量の点からの比較検討を行った。

## 2. 1 対訳文の字数

対訳文として、「Bertrand Russell's Best」(神山、1961) から最初の5編を選んで字数の比較を行った。日本語については漢字、仮名、空白の数を、英語については文字数と空白数に横原(1991)の値を用いた。日本語には濁点「・」半濁点「゜」音引「一」の数と漢字の画数を加えた。表2はこれらの字数の対訳文5編の合計の値である。

表2 対訳文の字数と漢字画数

英語字数			日本語字数					漢字 画数
英字	空白	計	漢字	仮名	空白	濁半濁	一	
3129	698	3827	517	1249	101	115	6	4497

## 2. 2 文字の画数

画数による比較をするために、まず文字の画数について考察した。

### (1) アルファベットの画数

文字は縦、横、斜め、曲線など線の要素の組合せで構成されている。実際に書くときには「画」の組合せで文字を組立てる。

英語では26文字と空白を直線的・一次的に組合せて単語を作り、その単語の組合せによって文章を作って情報を表現している。26個の各文字を早く正確に区別するための工夫が必要とされる。アルファベットの文字では簡単に早く書け、しかも区別が容易という条件を満たすように縦、横、斜め、曲線などの線素が効率よく組み合わせられている。

アルファベット26文字の画数の合計は、Iを1画とすると62画で1文字当りの単純平均画数は2.38画である。26文字の画数の出現確率(佐藤、1982)を考慮して1文字当りの平均画数を求めると、空白を含まない場合は2.37画、空白を含んだ場合には1.93画となる。また対訳文の最初の短編①の出現頻度の値を用いると、空白を含まない場合2.35画、空白を含む場合1.89画となり、佐藤(1982)の頻度のデータを用いた場合と殆ど同じ値となる。

表3 アルファベット(英語)の平均画数

	空白含まない	空白含む
佐藤(1982)の頻度による	2.37	1.93
対訳文①の頻度による	2.35	1.89
26文字の単純平均	(62÷26=)2.38	——

(注) Iは1画とした。

(2) 片仮名と平仮名の画数

平仮名の画数は46文字で合計104画であり、片仮名の画数は46文字で合計108画である。平仮名は文字を構成する線長は片仮名より長く、線の組合せも片仮名より複雑である。これが日本語の柔らかさの基本であろう。本稿では情報伝達の効率上では平仮名も片仮名も同じであるとし、仮名の画数として片仮名の画数の値を用いる。また濁点半濁点音引については、「゛」「゜」「一」を1文字とし各々2画、1画、1画とした。「゛」「゜」「一」を加えた片仮名の画数の合計は112画、平仮名の画数合計は108である。

さらに、片仮名の画を折れ曲がりで分けた線素に分解すると線素数合計は濁点、半濁点、音引を加えない場合133、加えた場合137となる。

漢字仮名混じり文での仮名の出現頻度に対訳文のデータを用いると、片仮名の平均画数は「゛」「゜」「一」を含んだ49文字で、空白を含まない場合2.20画、空白を含んだ場合2.05画となる。また片仮名を縦、横、斜めの線素に分解すると、平均線素数は空白を含まない場合2.49線素、空白を含む場合2.32線素となる。

表4 仮名の平均画数・平均線素数

対訳文	空白含まない				空白含む	
	片仮名画数	頻度	Σ [画数×頻度]	平均画数	頻度	平均画数
	1370	1370	3012	2.20	1471	2.05
	片仮名線素数	頻度	Σ [線素×頻度]	平均線素数	頻度	平均線素数
	1370	1370	3417	2.49	1471	2.32
	平仮名画数	頻度	Σ [画数×頻度]	平均画数	頻度	平均画数
	1370	1370	2924	2.13	1471	1.99
片仮名単純平均画数			$108 \div 46 = 2.35$	$112 \div 49 = 2.29$ : 含	゛ ゜ 一	
片仮名単純平均線素数			$133 \div 46 = 2.89$	$137 \div 49 = 2.80$ : 含	゛ ゜ 一	

(3) アルファベットと片仮名の画数からの考察

文字が縦、横、斜め（曲線）の4種類の線素から構成されていると考えると、

$$\text{アルファベット} : 26 = 4^{2.35} \quad \text{片仮名} : 49 = 4^{2.81}$$

となり、各々の計算上の平均線素数は2.35と2.81となる。これらの値はアルファベットの単純平均画数2.38、片仮名の単純平均線素数2.80とほぼ同じ値である。このことから、文字は縦、横、斜め（曲線）の4種の線素の組合せによって構成されそれによって各文字間の区別をしていると考えられる。

片仮名46文字の単純平均画数2.35はアルファベットの単純平均画数2.38に近い。しかし、片仮名の線素数の単純合計は137で画数の単純合計112より22%多い。

表音文字26文字および49文字が等確率で出現する時の平均情報量は

$$\log_2 26 = 4.70 \text{ビット}$$

$$\log_2 49 = 5.62 \text{ビット}$$

である。空白を含めず文字が等確率で出現する時の平均情報量は、仮名（49文字）がアルファベット（26文字）より20%多い。

これらのことから仮名では平均情報量は基本の線素を4種と考えると画数よりも線素数に関係があることになる。

また、片仮名の画の構成要素を5種類と仮定すると

$$49 = 5^{2.42} \quad , \quad 46 = 5^{2.38}$$

で、計算上の画数が2.40前後となり単純平均画数2.35と同じ程度となる。このことから、片仮名はアルファベットよりも画の基本要素が多いと考えてよい。アルファベットでは文字を区別する線素が画とほぼ一致する。アルファベットの基本文字数から計算される平均情報量は線素数または画数に依存する。

仮名は平均情報量はアルファベットより高く、画数はアルファベットと同じ程度である。仮名は基本文字数の増加により一文字あたりの線素数を増やし平均情報量を増加させるが、画の基本要素を増やすことによって画数をアルファベットと同じ程度に抑えている。

アルファベットと片仮名の画数についてまとめると次のようになる。

- ①文字は縦、横、斜め（曲線）の線素の組合せと考えることができる。  
一文字当りの線素数は言語の基本文字数によって決まり、一文字あたりの線素数は平均情報量に反映される。
- ②英語のアルファベットでは画は線素とほぼ一致する。アルファベットは4種の画の組合せと考えてもよい。
- ③日本語の片仮名の画の基本要素は線素の種類より多いと考えられる。そのため片仮名では線素数より少ない画数で大きな情報量を確保できる。仮名では1画の持つ情報量がアルファベットの1画の情報量より大きい。仮名の基本文字数がアルファベットの基本文字数より多いにもかかわらず、仮名とアルファベットの画数はほぼ同じである。

#### （4）漢字の画数

漢字の画の役割も片仮名の場合と同じと考えられる。漢字も画の組合せで文字を区別している。画は片仮名と同じく線素の組合せと考えられるので、画で漢字の（平均）情報量を表すことができることになる。

「漢字は空白の効果を持つ（漢字効果）」（横原、1991）ということからも、また漢字は多くの画からなって単独または熟語として英単語と等しい役割をすることからも、漢字は英語の単語間の空白の効果を含んでいると言える。

日本語では仮名も漢字もアルファベットのように空白の数ではなく、画の基本構成要素

素の種類と画数によって平均情報量を確保しているのである。

### 2.3 対訳文の総画数と情報量

前節の検討で、仮名とアルファベットの一字あたりの画数は同じ程度であるが、線素数は仮名の方が多く、したがって平均情報量は仮名の方が高いことがわかった。一方、文章中の空白の割合はより英文の方が高い（同書）。

異言語間の文字による情報伝達の比較では、個々の単語や文字による比較だけでなく、文章全体の比較も行うことが必要である。

対訳文の総情報量は日本語と英語で等しいことが分かっている（同書）ので、ここでは同じ内容を表す対訳文全体の比較を画数の点から行う。

#### (1) 総画数

同じ内容の情報を同じ程度の表現力で表していると思われる対訳文について、文章全体の文字の総画数を比較した。

英語の文章については、アルファベット一字当りの平均画数2.37を全字数に掛けることによって全体の総画数を求めた。日本語については平仮名と片仮名の字数の合計の総画数を片仮名画数を適用することによって求めた。日本語の文章全体の画数は仮名の総画数と漢字画数を合計した。

対訳文5編の合計では、英語の総画数は7416画、日本語の総画数は7509画で、両方の値はほぼ同じとなった。

表5 対訳文の総画数比較

対訳文	英語画数	日本語画数		
	英字数×2.37	漢字画数	仮名画数	総画数
①	1901	874	755	1629
②	1398	896	510	1406
③	1173	816	579	1395
④	1394	769	586	1355
⑤	1550	1142	582	1724
合計	7416	4497	3012	7509

#### (2) 画数と情報量

画数より仮名と漢字の情報量を求める。

アルファベットの平均情報量を1.0ビット総情報量を3827ビットとすると、アルファベット1画当りの情報量は空白の出現頻度を考慮すると0.46ビットとなる。仮名の総情報量は1568ビット、仮名と空白を合わせた総情報量は1684ビットとなる。この時、空白を含め

た時の仮名1文字当りの平均情報量は1.25ビットとなる。

漢字の画数から漢字全体の情報量を求めると、

$$0.46 \times 4497 \times 2.49 \div 2.20 = 2341 \text{ビット}$$

となる。仮名の総情報量1684ビットと漢字の総情報量2341ビットを加えると、3925ビットとなり次の式が成り立つ。

英語の総情報量  $\div$  仮名の総情報量 + 漢字の総情報量

(空白を含む) (空白を含む)

この時の漢字1字の平均情報量は4.53ビット、1画当りの平均情報量は0.52ビットとなる。漢字1字の平均情報量4.53ビットは実際の多重マルコフ情報源の場合の値である。漢字が等確率で出現するとした場合の平均情報量は、基本漢字数を8192とすると13.0ビットとなる。漢字が等確率で出現する時と実際との比(13.0  $\div$  4.53) 2.87は仮名の比(5.64  $\div$  1.25 = 4.51) やアルファベットの比(4.76  $\div$  1.0 = 4.76) より小さい。これは漢字の使用頻度の均等性やマルコフ連鎖が短いことや「漢字効果」(横原, 1991) による。

### 3. 結論

象形文字に分類される漢字と対応する英単語の比較から、基本的な「実体」を表す単語については漢字の画数と英単語の字数に1対1に近い対応があることがわかった。

文字で表す情報の少ない時代にできた象形文字に分類できる漢字では「画」1つが現在のアルファベット1文字と同じ役割を果たしているのである。

#### 〈基本漢字と英単語との関係〉

漢字のうち基本的な「実体」を表し象形文字に分類できる漢字の画数は、対応する英語の単語の文字数に近い。実際に使用されている時の平均情報量を考えると、基本的な漢字の1画がおおよそ1.0ビットの情報量を持つと言える。

実際に使われている漢字数はこの基本的な漢字数よりはるかに多い。漢字全体では平均的に、漢字の画数から情報量を次のように推定できる。

#### 〈漢字の画数と情報量〉

漢字の画数に0.52ビットを掛ければ漢字の情報量のおおよその値が求まる。

対訳文全体の画数の比較から、同じ内容を表す英語の画数と日本語の画数は同じであることがわかった。



〈日本語と英語の間における総画数一定の関係〉

日本語と英語の間では、同じ内容の情報を伝える文章全体の文字の総画数は等しい。

このことは両言語の文章表現力、言いかえれば情報表現の効率が同じであることが前提である。この関係は日本語と英語の間だけでなく、長い年月をかけて発達し成熟し終わっているどの二言語間についても、さらに時代の隔たりのある同じ言語の間でも成り立つのではなかろうか。

〈異文字間における総画数一定の関係…仮説〉

同じ内容の情報を伝える文章全体の文字の総画数は、言語が異なっても等しい。

〈時代差のある同じ言語間における総画数一定の関係…仮説〉

同じ内容の情報を伝える文章全体の文字の総画数は、時代差のある同じ言語間で等しい。

総情報量についても同じことが言える。

〈異文字間における総情報量一定の関係…仮説〉

同じ内容の情報を伝える文章全体の文字の総情報量は、言語が異なっても等しい。

〈時代差のある同じ言語間における総情報量一定の関係…仮説〉

同じ内容の情報を伝える文章全体の文字の総情報量は、時代差のある同じ言語間で等しい。

本稿の画数に関する議論は、文字の情報処理機械が発達する以前に長い年月をかけて成熟してきた言語と文字について言えるのである。画数と情報量の観点から、日本語と英語以外の言語を含めて比較することが今後の課題である。機械による文字の情報処理化がより普及進展した時代の文字についての議論の進展が待たれる。

参考文献

- 神山正治 1961 『Bertrand Russell's Best』 金星堂  
横原恭士 1991 「日本語における漢字の役割について」『相愛大学研究論集』第8号  
佐藤憲市 1982 『情報科学』八千代出版