

テーマ	確率で英文を作ろう				
概要	英文を文字ではなく、アルファベットという記号が並んでいるという前提で、その頻出度を利用して機械的に並べながら、英文にならないかを手作業で考える。その中で、マルコフ過程の意味を理解し、応用する。また、さらに手作業での流れをコンピュータを活用してできないかを考える。				
ねらい	問題解決においては、解決の手順と用いる手段の違いが結果に影響を与えること及びコンピュータの適切な活用が有効であることを理解させる。				
関連する主な科目・項目	情報B「問題解決における手順とコンピュータの活用」	難易度	① 易 普・難	想定時間数	約2時間
準備するもの	適当な英文, Excel	作成者	亀山 弘		

1 学習の展開

(1) 導入

英語の英文は、26文字のアルファベットと、単語と単語の間のスペースとで作られている。つまり、27の記号で構成されていると考えてよい。その27の記号がどのような頻度で使われているのかを確率で表したものが、次の表である。

表

文字	頻度	文字	頻度
A	0.0668	O	0.0654
B	0.0118	P	0.0162
C	0.0226	Q	0.0010
D	0.0310	R	0.0559
E	0.1073	S	0.0499
F	0.0239	T	0.0856
G	0.0163	U	0.0201
H	0.0431	V	0.0075
I	0.0519	W	0.0126
J	0.0011	X	0.0014
K	0.0034	Y	0.0162
L	0.0278	Z	0.0006
M	0.0208	スペース	0.1817
N	0.0581		

そこで、その確率に従うように記号を選び出し、英文を作ってみよう。
例文として用意した平凡な英文で、具体的に説明してみよう。

<例文>

Early in life I learned what it means to be unable to do what others can do.
I was born with stumps where legs ought to be.
For my first six years a hospital was my home.
After that I had to wear special boots.
My legs were so short that my arms almost touched the ground.
Some of the boys near my home laughed at me and called me names, and so I learned to fight.
When I was twentyfive years old I was only three feet eight inches tall.

Then a wonderful thing happened.

A doctor named Robert Yanover took me to see a friend who was clever at making things with his hands.

He made a pair of aluminum legs for me.

When they put the aluminum legs on me, I was five feet eight inches tall.

Suddenly everyone realized that I was a man. (以下省略)

< 第一段階 >

この英文は、27 の記号の発生確率に従って作り出されているはずなので、これをバラバラにして使えば27 の記号の乱数として使えるはずである。そこで、始めから10 番目ごとに文字を抜き取って並べてみる。単語間のスペースも1つの記号として数えることを忘れないでほしい。

L R I O E A C A T W B F Y S F I E L T M T E O A M L E I T H T T Y E N N I E
R B E F R G I N E F M E L E H T N N D A O V D A R H H R M E E L H E E
T V E N U E

この文章は、27 の記号の使用確率については、英語をまねしたのに、あまり英語らしくない。それもそのはずである。英語には、文字が現れる確率の他に、もう少し英語としての癖があるからである。例えば、Q という文字の次には必ずU が来るし、B の次にはE が並びやすく、T はめったに来ないというような癖である。「ある文字が来た時、その次にある文字が現れる確率」を考えなければならない。このように、現在の状態が1 回前の状態で決まるような確率の過程を『単純マルコフ過程』と呼んでいる。

< 第二段階 >

そこで、単純マルコフ過程と考えて、英文を作ってみよう。

先ほどの英文の10 番目の文字L から始める。英文の始めからL を捜していくと、最初の単語の中に見つけることができる。L の次はY である。そこで、

L Y

と記録する。そして、ここのL Y という文字は、二度と使わないようにし、鉛筆で印を付けておく。次に、Y の文字を捜す。最初の単語のY はもう使ったので、これはとばしていく。次を見ていくとmy という単語があり、Y が見つかります。Y の次はスペースである。そこで、次には英文の最初からスペースを捜して、その次のI を採用する。こうして、ある文字が来たとき、その次の文字が何であるのかの確率を英語に等しくなるようにして英文を作ってみると、次のようになる。

L Y I F O L E A R S I T O W H E D O R E A T M P I U G S D C I T U C A B
E W A N D S T W A N E L O T O S I M Y E A R H E A L T W E G R E M
Y S H A F

まだ、あまり英語らしいとは言えないが、それでも、heal や my のような見慣れた単語も現れてきて、先ほどより英語に近い気がする。単純マルコフ過程を、さらに発展させ、現在の状態が1 回前と2 回前の状態で決まるような確率の過程を『2 重マルコフ過程』と呼んでいる。

< 第三段階 >

さらに、英文を2 重マルコフ過程として考えてみよう。つまり、文字の現れる確率は、その前の文字と、さらにもう1つ前の文字とによって決まると考える。今作った英文の最初の2文字L Y から始める。英文の頭からL Y を捜していくと、最初の単語がすぐに見つかる。前に作った英文のL Y は、ここのL とずっと後のY とで作ったので、ここのL Y とは関係がない。L Y の次のスペースだから、スペースを で表すことにし、まず

LY

を採用する。次は、Y の後に来る文字を見つける。ずっと見ていくとY Fがある。さらに、Fの後の文字を、一度使った文字は使わないようにして搜すと、 fight という単語があり、I を選ぶことができる。このようにして、2重マルコフ過程として考えて作られた英文は次のとおりである。

LY FIR TO DO WHAT MY HOMEARN LE INCHERE USED
TO BOOKE FE ARS WAS CAND ANORT ST ITHE LEGS
OF TO ATENT TWO SINCERD

今度は、知っている単語がたくさん並んでいる。いかにも、英文らしくなってきた。

この調子で、文字の出現確率を3重マルコフ過程や4重マルコフ過程として考えて英文を作っていくと、ますます英文らしくなってくる。このように、人間の意志を入れずに文章を作り出すことは、ドイツ語でも日本語でもできる。人間の意志によらず作られた文章に、もし重大な内容があったとしたら、それは私たちに対して、どのような意味を持っているのかを考えると面白いですね。

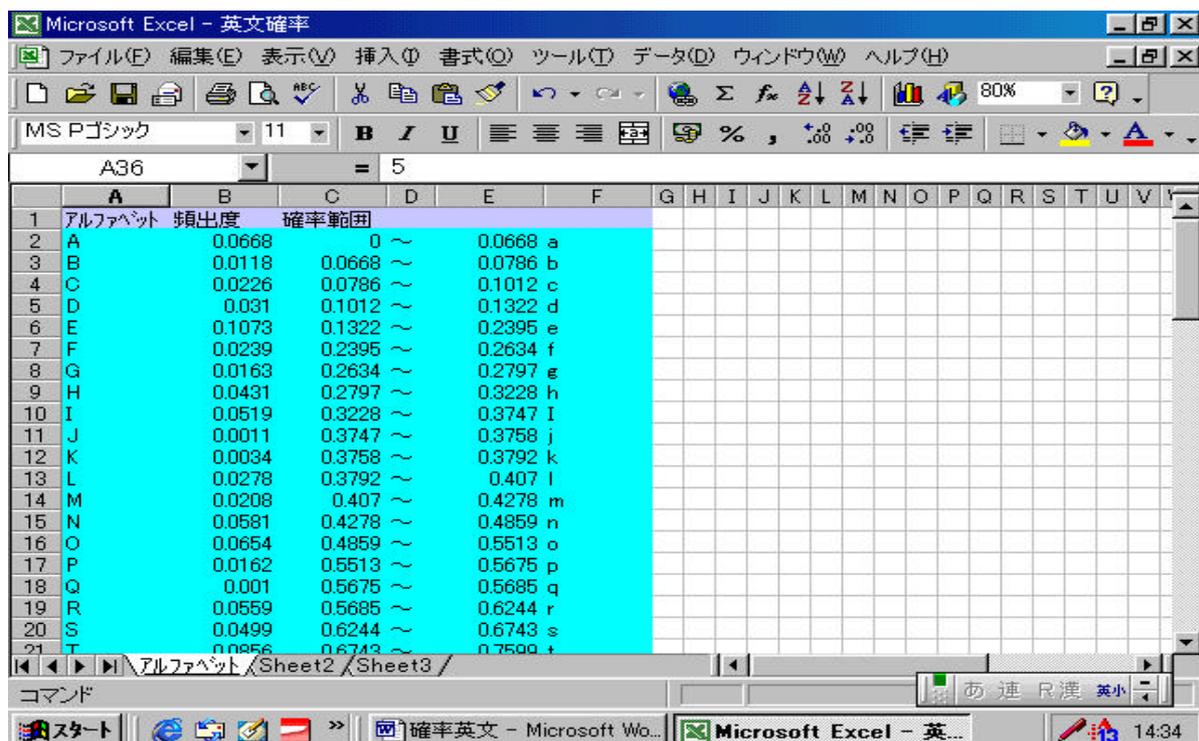
(2) 問題提起

各自が持参した適当な英文で、同じように調べてみよう。

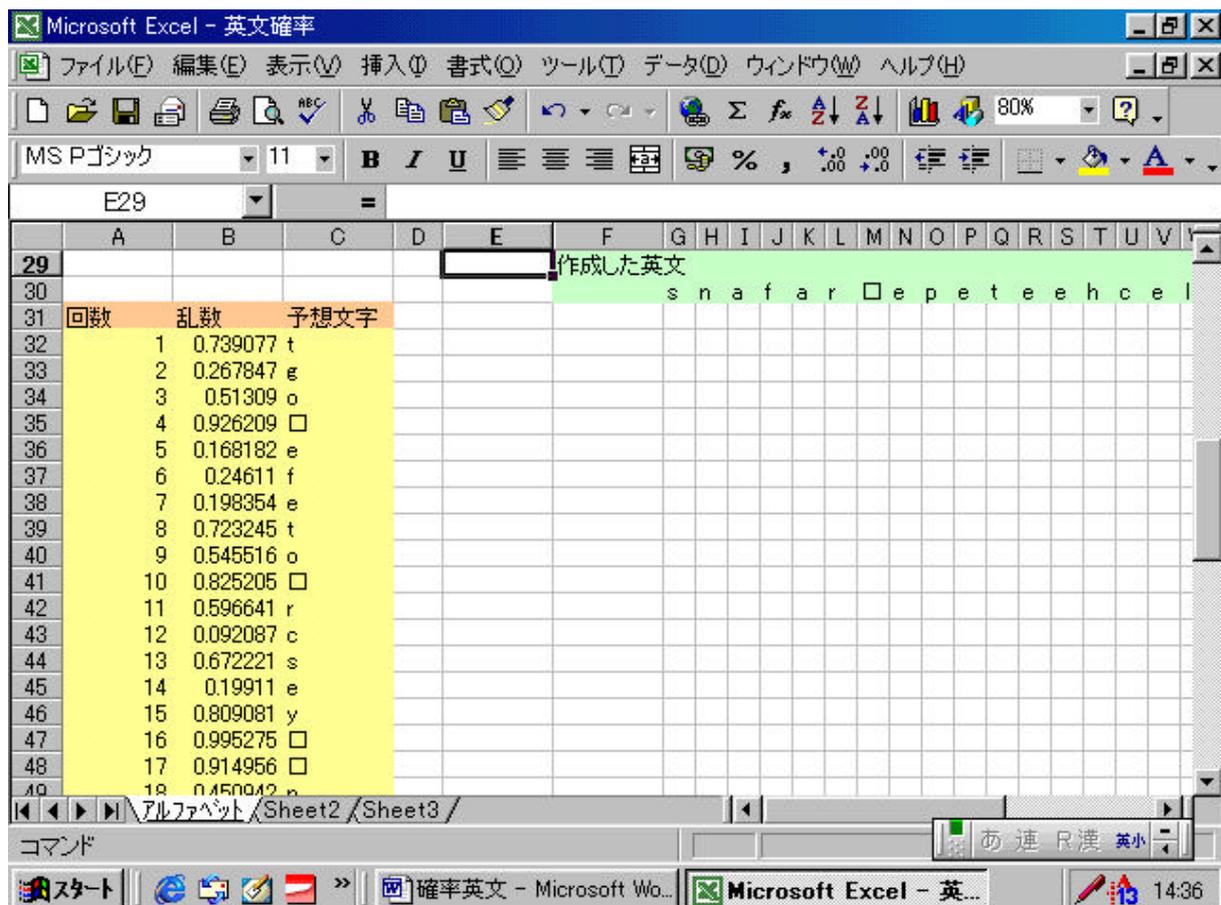
2 発展的な学習

(1) 乱数を用いた英文

表 の確率で英文が構成されているという前提で、乱数を発生させアルファベットを並べてみよう。Excel を用いて実験してみよう。



1	アルファベット	頻出度	確率範囲	
2	A	0.0668	0 ~ 0.0668	a
3	B	0.0118	0.0668 ~ 0.0786	b
4	C	0.0226	0.0786 ~ 0.1012	c
5	D	0.031	0.1012 ~ 0.1322	d
6	E	0.1073	0.1322 ~ 0.2395	e
7	F	0.0239	0.2395 ~ 0.2634	f
8	G	0.0163	0.2634 ~ 0.2797	g
9	H	0.0431	0.2797 ~ 0.3228	h
10	I	0.0519	0.3228 ~ 0.3747	i
11	J	0.0011	0.3747 ~ 0.3758	j
12	K	0.0034	0.3758 ~ 0.3792	k
13	L	0.0278	0.3792 ~ 0.407	l
14	M	0.0208	0.407 ~ 0.4278	m
15	N	0.0581	0.4278 ~ 0.4859	n
16	O	0.0654	0.4859 ~ 0.5513	o
17	P	0.0162	0.5513 ~ 0.5675	p
18	Q	0.001	0.5675 ~ 0.5685	q
19	R	0.0559	0.5685 ~ 0.6244	r
20	S	0.0499	0.6244 ~ 0.6743	s
21	T	0.0856	0.6743 ~ 0.7599	t



(2) マルコフ過程

上の Excel では、ただ乱数を発生させ、アルファベットを並べただけである。マルコフ過程に従って同じ演算を行うためには、2つのアルファベットの文字が並ぶ頻度を表す確率が必要になる。2重マルコフ過程、3重マルコフ過程と調べていくためにも、アルファベットが3つや4つの文字が並ぶ頻度を表す確率が必要である。気が遠くなるようなことであるが、コンピュータを活用すれば、なんとかならないだろうか。考えてみよう。

3 参考文献

大村 平 「確率のはなし - 基礎・応用・娯楽 - 」 日科技連出版社 1993