

漢文訓読の数学モデル

— 返り点と豎点が生み出す漢字列の総数 —

島野 達雄 (関西学院大学)

1. 漢文訓読の基本構造

1.1 英文和訳と漢文訓読

明治の初め, What are you writing? という英文を, あなたは(you)何を(What)書いて(writing)ござるか(are)と訳し, 漢文のように返り点や豎点^{たててん}をつけようとしてうまくゆかず, 仕方なく What₂are₄you₁writing₃? のように番号をつけたという逸話がある.

これは What(A) are(B) you(C) writing(D)? を, あなたは(C)何を(A)書いて(D)ござるか(B), つまり, ABCD を CADB の語順にする, 返り点や豎点のつけ方がないことを示している. ここで, 返り点とはレ点, 一二点, 上下点などを指し, 豎点とは漢字連結記号—を指す. 実際に書きだすと, 返り点だけからは, ABCD から

ABCD=ABCD, ABC_∨D=ABDC, AB_∨CD=ACBD, AB_∨CD_∨D=ADCB,
A_∨BCD=BACD, A_∨BC_∨D=BADC, A_∨BC_∨D=BCAD, A_∨BCD_∨=BCDA, A_∨BC_∨D_∨=BDCA,
A_∨B_∨CD=CBAD, A_∨B_∨CD_∨=CBDA, A_∨B_∨CD_∨=CDBA, A_∨B_∨C_∨D=DCBA

の 14 通りの漢字列が生み出される. 豎点を使うと, この 14 通りに,

AB—_∨CD_∨=ADBC, A—_∨BC_∨D=CABD, A—_∨BCD_∨=CDAB,
A—B—_∨CD_∨=DABC, A_∨B—_∨CD_∨=DBCA, A—_∨BC_∨D_∨=DCAB

の 6 通りが加わり, 合計 20 通りになる. ちなみに並べかえの総数は $4! = 24$ 通りある. 本稿は, 一般に n 字の漢字列から, 返り点と豎点によっていくつの漢字列が生み出されるか, とくに豎点を使わず返り点だけから生み出される漢字列の総数を数学的に求め, 漢文と日本語の親近性を量的に示すことを目的にしている.

1.2 返読の基本構造

漢文訓読には, 「返って読まれるのは, 漢字 1 字または豎点で結ばれた 2 字以上の漢字列」という特徴がある. たとえば, 先に読む部分を <> で囲むと, レ点を使う例文では, 不_∨惑は不<惑>, 不_∨能_∨闘_∨力は不<能<闘<力>>>, 一二点とレ点では, 偶々有_∨名酒_∨無_∨夕不_∨飲は偶々有<名酒>無<夕不<飲>>, 豎点があるときは, 攘—_∨除姦凶—以復—_∨興漢室—は攘—_∨除<姦凶>以復—_∨興<漢室>, のように表記できる.

つまり先に読む部分を α として $\langle \alpha \rangle$ のように $\langle \rangle$ で囲み、後で読む漢字1字または豎点で結ばれた2字以上の漢字列を β としたとき、 $\beta \langle \alpha \rangle$ が返読の基本的な構造になる。

一二点のなかにレ点があるように、 α のなかに別の返読構造が何重にも入れ子になっていてもよいし、複数の返読構造が並置されていてもよい。むろん文全体として、このような返読構造がいくつも並置されていてもよい。

2. 数学的形式化

2.1 漢文言語・接続・返接とカッコつき表示(I)

数学的形式化の第一歩として、漢文言語および2つの二項演算を次のように定義する。すべての漢字(無定義用語)の集合を $\Sigma = \{\sigma_1, \sigma_2, \dots, \sigma_n\}$ とする。

Σ の元によるすべての順列 $\Sigma^* = \{\lambda, \sigma_1, \dots, \sigma_n, \sigma_1\sigma_1, \sigma_1\sigma_2, \dots, \sigma_1\sigma_n, \sigma_2\sigma_1, \dots\}$ を漢文言語とよび、 Σ^* の元を漢字列とよぶ。ここで λ は空順列を指し、空の漢字列とよぶ。

Σ^* の元 u, v に対して、 u, v を並べる二項演算、 $u \circ v = uv$ を接続とよぶ。

同様に、 u, v の順序を入れかえる二項演算、 $u \langle v \rangle = vu$ を返接とよぶ。この返接演算において、 u を構成する漢字が、たとえば $u = \sigma_3\sigma_1\sigma_3\sigma_2$ であるとき、連結記号 $-$ を用いて $u \langle v \rangle = \sigma_3-\sigma_1-\sigma_3-\sigma_2 \langle v \rangle$ と表記する。

以下、接続の記号 \circ を省略し、接続に対して返接が優先するとする。

漢文言語 Σ^* の λ 以外の漢字列に、0個以上、返接の記号 $\langle \rangle$ を付け加え、 \langle に先行する漢字列が2字以上であればそれらを連結記号で結んだものをカッコつき表示(I)とよび、次の①~③により再帰的に定義する。カッコつき表示(II)については後述する。

①漢字1字はカッコつき表示である。

②カッコつき表示とカッコつき表示の接続はカッコつき表示である。

③ β を漢字1字または連結記号で結ばれた2字以上の漢字、 α をカッコつき表示とするとき、 β と α の返接 $\beta \langle \alpha \rangle$ はカッコつき表示である。

カッコつき表示に対して、返接演算をおこない、連結記号を取り除いて漢字だけを並べた列を、カッコつき表示の展開、カッコをはずした結果などとよぶ。

2.2 訓読の木

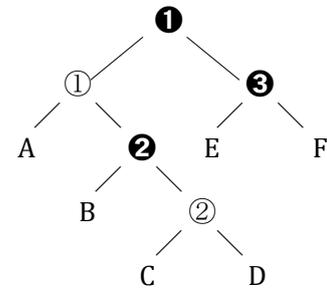
漢字列が与えられたとき、「接続や返接によって、どんな漢字列が得られるか」を示すグラフを描くことができる。このグラフを訓読の木とよぶ。定義は後述する。

訓読の木は、① Σ の要素である漢字、②接続ノード(●であらわす)、③返接ノード(○であらわす)で構成される。2つのノードとも左辺と右辺をもち、これらの先にはノードや漢字をおく。あるノードに対して、左辺や右辺の先におかれたノードや漢字をこのノードの下位とよぶ。このノード自体は上位にあたる。

訓読の木の読み方を示しておこう。ノードの左右両辺は、往路と復路の2回通過する。まず最上位のノードから始める。最上位が接続ノードであれば左辺に進み、返接ノード

であれば右辺に進む。以降、往路でも復路でも接続ノードに出会えば、進行方向に向かって右側の辺を進み、返接ノードに出会えば、進行方向に向かって左側の辺を進む。こうして最下位の漢字に到達すると、その漢字を読んだあとUターンする。このようにして、すべての漢字を読み、最上位のノードまでもどる。

カッコつき表示 $A\langle BC\langle D\rangle\rangle EF$ (返り点では $A_BC_ \downarrow DEF$) の訓読の木は、接続ノード●と返接ノード○に番号をつけて、右図のようになる。上のルールで読むと、①→①→②→B→②→②→D→②→C→②→②→①→A→①→①→③→E→③→F→③→①となり、確かに $A\langle BC\langle D\rangle\rangle EF$ を展開した $BCDAEF$ の順に漢字を読んでいる。



2.3 順列行列による表現

カッコつき表示とその展開、 $A\langle BCD\rangle=BCDA$ (返り点では $A_BCD_$)は、 $ABCD$ を $BCDA$ に並べかえる置換とも解釈できる。このような置換は、各行各列に 1 っだけ 1 があり、他の成分がすべて 0 の順列行列で表現できる。上の例では、

$$(A \ B \ C \ D) \begin{pmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = (B \ C \ D \ A) \text{ となる.}$$

以下、行列とその()を省いたものとは同一視し、成分の 0 を省略する。

i 次の順列行列 A_i と、 $n-i$ 次の順列行列 A_{n-i} に対し、

$$\begin{pmatrix} A_i & \\ & A_{n-i} \end{pmatrix} \text{ を } A_i \text{ と } A_{n-i} \text{ の対角配置, } \begin{pmatrix} & A_i \\ A_{n-i} & \end{pmatrix} \text{ を } A_i \text{ と } A_{n-i} \text{ の逆対角配置とよぶ.}$$

訓読の木と同様、接続と返接に対応する順列行列を訓読行列とよぶ。対角配置や逆対角配置を用いた、訓読行列の再帰的定義は次節で述べる。

2.4 カッコつき表示(II)・訓読の木・訓読行列の定義

カッコつき表示(I)の定義③では、返接の β 項を「漢字 1 字または連結記号で結ばれた 2 字以上の漢字」と、連結記号を用いて定義したが、この β 項自体を連結項として再帰的に定義できる。訓読の木では、これを連結枝として定義する。順列行列では、1 行 1 列の 1 次の単位行列が「漢字 1 字」に対応し、 k 次の単位行列が連結項に対応する。

連結項の定義	訓読の木の連結枝の定義	訓読行列の場合
①漢字 1 字は連結項である。	①漢字 1 字は連結枝である。	※1 次の単位行列が「漢字 1 字」に対応し、 k 次($k \geq 1$)の単位行列が「連結項」や「連結枝」に対応する。
②連結項と漢字 1 字の接続は連結項である。この接続を連結記号—であらわす。	②左辺が連結枝、右辺が漢字 1 字の接続ノードは連結枝である。	

この連結項、連結枝、 k 次の単位行列を用いて、カッコつき表示(II)、訓読の木、訓読行

列を次のように定義する.

カッコつき表示(II)の定義	訓読の木の定義	訓読行列の定義
①漢字 1 字はカッコつき表示である.	①漢字 1 字は訓読の木である.	①1 次の単位行列は訓読行列である.
②連結項とカッコつき表示の返接はカッコつき表示である.	②左辺が連結枝, 右辺が訓読の木の返接ノードは訓読の木である.	② k 次($k \geq 1$)の単位行列と訓読行列の逆対角配置は訓読行列である.
③漢字 1 字または②の条件を満たす返接と, カッコつき表示の接続はカッコつき表示である.	③左辺が漢字 1 字または②の条件を満たす返接ノード, 右辺が訓読の木の接続ノードは訓読の木である.	③1 次の単位行列または②の条件を満たす逆対角配置と, 訓読行列の対角配置は訓読行列である.

カッコつき表示(II)が(I)を満たすことは明らかで, 逆に(I)であれば, (I)②の接続の先行する項が(II)②の返接の項であれば(II)③を満たしているし, 先行する項が漢字に始まる接続であれば(II)③を満たしているので, (I)と(II)の定義は同等な定義になる.

これら 3 つの定義は, 次章の漢字列の数えあげに有効になる. 訓読行列の定義を使うと, 逆行列 (順行列列では必ず転置行列になる) が訓読行列か否かを容易に判定できる.

2.5 広義の返読構造モデル

なお, 漢文英訳, 英文和訳などでは漢文訓読より広義の返読構造が見られる. これまで述べてきたように, 通常の漢文訓読では, 返って読まれる語句が漢字 1 字または連結記号で結ばれた 2 字以上の漢字であるが, 漢文英訳, 英文和訳などでは, 返って読まれる語句のなかで返読が起こることがある.

広義の返読構造モデルは, 返接の β 項の違いによって, 次の 3 つに区分できる.

- ① β 項が漢字 1 字に限られている場合. 単返接モデルとよぶ.
- ② β 項が k 字($k \geq 1$)の漢字列の場合. k 連返接モデルまたは連返接モデルとよぶ.
- ③ β 項に{ }記号の使用を認め, β 項のなかに返接がある場合. 複返接モデルとよぶ.

これらの再帰的定義は, ①単返接モデルでは連結項, 連結枝を「漢字 1 字」に置きかえ, ②複返接モデルでは連結項, 連結枝を「左右両辺に訓読の木がある接続ノード」に置きかえれば可能であるが, たとえば訓読の木の一意性は保たれなくなる.

3. 訓読漸化式と一般解

3.1 連返接モデルの訓読漸化式

n 字の漢字列が与えられたとき, そこから生じるカッコつき表示の総数を a_n とする. a_0 は空の漢字列 λ の置換とみなし, $a_0 = 1$ と定義する. 漢字 1 字では, $a_1 = 1$ となる. $n \geq 2$ のとき, 最初の漢字が接続, 返接のどちらの項に属するかによって, a_n は次の 2 通りの, 場合の数の和になる.

①最初の漢字が接続の先行する項になっている場合. このときのカッコつき表示の数は, 最初の1字を除いた総数つまり a_{n-1} に等しい.

②最初の漢字が返接の β 項の一部である場合. 返接の β 項に k 字($1 \leq k \leq n-1$), α 項に i 字($1 \leq i \leq n-k$)あるとすると, 残りに $n-i-k$ 字あることになり, α 項から a_i , 残りから a_{n-i-k} が生み出されるので, その数は a_i と a_{n-i-k} の積の総和になる.

よって, ①②をあわせて,

$$a_n = a_{n-1} + \sum_{k=1}^{n-1} \sum_{i=1}^{n-k} a_i a_{n-i-k} = a_{n-1} + \sum_{i=1}^{n-1} a_i a_{n-i-1} + \sum_{i=1}^{n-2} a_i a_{n-i-2} + \cdots + \sum_{i=1}^{n-(n-1)} a_i a_{n-i-(n-1)}$$

$$= a_{n-1} + (a_1 a_{n-2} + a_2 a_{n-3} + \cdots + a_{n-1} a_0) + (a_1 a_{n-3} + a_2 a_{n-4} + \cdots + a_{n-2} a_0) + \cdots + a_1 a_0$$

を得る.

この再帰公式(訓読漸化式とよぶ)は, 次のように簡略化できる(大西正男).

$$a_{n+1} = 3a_n + a_{n-1} + \sum_{i=2}^{n-2} a_i a_{n-i} \quad (n \geq 4)$$

3.2 単返接モデルの訓読漸化式

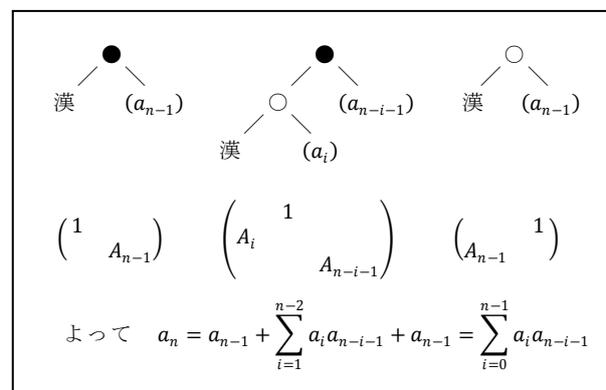
返接の基本構造 $\beta < \alpha >$ で, β 項が漢字1字だけの場合, すなわち単返接モデルを考えてみよう. このようなケースに限定してもよいと思われる根拠は, 豎点を使わずに返り点だけで, おおむね漢文の訓読がおこなえるからである. 患_レ所_一以_レ立_レ (立つをもってする所を患う)と読んで, 攘_一除_レ姦凶_一以_レ復_一興漢室_一を攘_レ除_レ姦凶_一以_レ復_一興_レ漢室_一 (姦凶を除き攘い, もって復漢室を興す)と読んで意味はさほど変わらない.

単返接モデルの訓読漸化式は, 連返接モデルの β 項を漢字1字に限定すればよい. つまり $k=1$ のときの総和だけにするとよい. すなわち

$$a_n = a_{n-1} + \sum_{i=1}^{n-1} a_i a_{n-i-1} = a_{n-1} a_0 + (a_1 a_{n-2} + a_2 a_{n-3} + \cdots + a_{n-1} a_0) = \sum_{i=0}^{n-1} a_i a_{n-i-1}$$

この式は, 訓読の木と訓読行列からも得られる(右図). 訓読の木は, 左の2つが最上位に接続ノードがあるときで, 右は最上位に返接ノードがあるときを示している. それぞれの右辺には, 下位につながる漢字列の総数を書いている.

3つの訓読行列の中央は, 1次の単位行列と A_i の逆対角配置, それと A_{n-i-1} の対角配置を示している.



3.3 単返接モデルの一般項と極限値の意味

以下, 母関数を用いて単返接モデルの訓読漸化式の一般項 a_n を求めよう.

$$g(x) = a_0x^0 + a_1x^1 + a_2x^2 + \dots + a_kx^k + \dots \quad \text{とおくと,}$$

$$\begin{aligned} \{g(x)\}^2 &= a_0^2 + (a_0a_1 + a_1a_0)x + \dots + (a_0a_{k-1} + a_1a_{k-2} + \dots + a_{k-2}a_1 + a_{k-1}a_0)x^{k-1} + \dots \\ &= a_1 + a_2x + \dots + a_kx^{k-1} + \dots \quad (\because a_0 = a_1 = 1, \quad a_n = \sum_{i=0}^{n-1} a_i a_{n-i-1}) \end{aligned}$$

これより, $x\{g(x)\}^2 + a_0 = g(x)$ を得る.

$g(x)$ を g と書いて2次方程式を解くと, $g = \frac{1 \pm \sqrt{1-4x}}{2x}$ を得るが, プラスの解は, 拡張

二項定理で $\sqrt{1-4x}$ を展開すると, $g(x)$ の初項が定数項にならないので不適.

$$\begin{aligned} (1-4x)^{\frac{1}{2}} &= 1 + \sum_{k=1}^{\infty} \binom{\frac{1}{2}}{k} (-4x)^k = 1 + \sum_{k=1}^{\infty} \frac{\frac{1}{2}(-\frac{1}{2})(-\frac{3}{2})\dots(\frac{1}{2}-k+1)}{1 \cdot 2 \cdot 3 \cdot \dots \cdot k} (-4x)^k \\ &= 1 - 2x + \sum_{k=2}^{\infty} \frac{2^k(-1)1 \cdot 3 \cdot 5 \cdot \dots \cdot (2k-3)}{1 \cdot 2 \cdot 3 \cdot \dots \cdot k} x^k = 1 - 2x - 2 \sum_{k=2}^{\infty} \frac{2 \cdot 6 \cdot 10 \cdot \dots \cdot (4k-6)}{2 \cdot 3 \cdot 4 \cdot \dots \cdot k} x^k \\ \therefore g(x) &= \frac{1}{2x} \left[1 - \left\{ 1 - 2x - 2 \sum_{k=2}^{\infty} \left(\frac{2}{2} \right) \left(\frac{6}{3} \right) \left(\frac{10}{4} \right) \dots \left(\frac{4k-6}{k} \right) x^k \right\} \right] \\ &= 1 + \sum_{k=2}^{\infty} \left(\frac{2}{2} \right) \left(\frac{6}{3} \right) \left(\frac{10}{4} \right) \dots \left(\frac{4k-6}{k} \right) x^{k-1} \\ &= 1 + \sum_{n=1}^{\infty} \left(\frac{2}{2} \right) \left(\frac{6}{3} \right) \left(\frac{10}{4} \right) \dots \left(\frac{4n-2}{n+1} \right) x^n \quad (k-1=n \text{とおいた}) \end{aligned}$$

ここで, a_n は x^n の係数だから,

$$a_n = \left(\frac{2}{2} \right) \left(\frac{6}{3} \right) \left(\frac{10}{4} \right) \dots \left(\frac{4n-2}{n+1} \right) \quad \text{という一般項を得る. この} a_n \text{は}$$

$$2 \cdot 6 \cdot 10 \cdot \dots \cdot (4n-2) = 1 \cdot 3 \cdot 5 \cdot \dots \cdot (2n-1) \cdot 2^n = \frac{1 \cdot 2 \cdot 3 \cdot \dots \cdot (2n-1)(2n)2^n}{2 \cdot 4 \cdot 6 \cdot \dots \cdot (2n)} = \frac{(2n)!}{n!} \quad \text{より}$$

$$a_n = \frac{(2n)!}{n!(n+1)!} = \frac{1}{n+1} \binom{2n}{n} \quad \text{とあらわすことができる. また,}$$

$$a_{n+1} = \frac{4n+2}{n+2} a_n \quad \text{となるので, } \lim_{n \rightarrow \infty} \frac{a_{n+1}}{a_n} = \lim_{n \rightarrow \infty} \frac{4n+2}{n+2} = 4 \quad \text{となる.}$$

この極限値の意味は, 次のように解釈することができる.

豎点を使わずに返り点だけで漢文訓読がおおむね可能なことは, すでに述べた.

このように豎点の使用を認めない場合, 十分に大きな漢字列では, 1字増えるごとに日本語への翻訳の可能性が4倍に増えるだけで, 可能性が極端に増えるわけではない.

以上のように, 意味を捨象した訓読のアルゴリズムの解析からだけでも, 漢文すなわち中国語と日本語は, 親近性が高いと言える. (了)