

漢文訓読の代数的構造ⁱ⁾

一 返り点・豎点が生み出す「返読」の数え上げ組合せ論一

島野達雄(関西学院大学共通教育センター)ⁱⁱ⁾

1. 漢文訓読とカッコ・豎点表示

漢文訓読とは、中国語の漢字列に返り点、豎点などを付けて日本語化する「翻訳」を意味する。歴史的産物のため、翻訳者によって日本語訳が異なるほか、返り点や豎点などの付け方も一様ではない。返り点や豎点は、「中国語から日本語へ」という 2 つの自然言語間の翻訳をあつかう「自然メタ言語」の一部分を構成している。

ここでは、個々の漢字の意味や役割などを無視し、返り点と豎点の付け方のパターンだけを考察の対象にする。たとえば、

不 _レ 如 _一 一見 _一	(百聞は) 一見に如かず	[比較・否定]
無 _レ 俾 _一 民憂 _一	民をして憂へしむる無かれ	[使役・命令]
宜 _レ 在 _一 高位 _一	(惟仁者のみ) 宜しく高位に在るべし	[再読]

の 3 つは同じ返り点のパターンを持っているとみなす。また、中国語の「有楚大夫於此」に対して、「有_レ楚_一大_一夫_一於此_一」「有_二楚_一大_一夫_一於此_一」「有_レ楚_一大_一夫_一於此_一」などの返り点・豎点の付け方があるが、これらはすべて「此に楚の大夫有り」と訓読するので、同じパターンと見なす。

さて、(I) n 字の中国語の漢字列が与えられたとき、返り点・豎点の付け方のパターンは何通りあるだろうか。

漢文訓読における「返読」の特徴の一つに、「返って読まれるのは漢字 1 字または豎点でつながれた 2 字以上の漢字」がある。この返って読まれる部分を β 、 β より先に読む部分を α とすると、返読は $\beta < \alpha$ とあらわせる。説明の便宜上、返り点・豎点をこのようにカッコの組($< >$)と豎点($-$)に書きかえ、「カッコ・豎点表示」とよぶ。

[表 1]は、代表的な返り点・豎点つき漢文がカッコ・豎点表示に書きかえられることを示している。

このカッコ・豎点表示を観察すると、もとの n 字の中国語の漢字列 $a = a_1 a_2 \cdots a_n$ の最初の 2 字の関係は、(1) $a_1 a_2$ 型(①②⑦⑨)、(2) $a_1 < a_2$ 型(③⑤⑥⑩⑪)、(3) $a_1 - a_2$ 型(④⑧)の 3 通りしかないことがわかる。

また、カッコの組は、一二点の中にレ点があるとき(⑤)や上下点の中に一二点があるとき(⑥)などで入れ子になるほか、一二三点(⑦⑧)、レ点の連続(⑨)、レ点と一二点の複合(⑪)などでも入れ子になっている。

さらに、 $>$ のすぐ右側は、「 $>$ 」(⑦の子 $>>$ 、⑨の簪 $>>>$ など)、「漢字 1 字」(③時 $>$ の花、⑤佩 $>$ の玉、⑥里 $>$ の称、称 $>$ の也)、または「空(empty)」(②の月 $>$ 、③の涙 $>$ など)になっている。

ⁱ⁾ Douglas Rogers, 齋藤明両氏の協力により 2012 年 12 月の応用数学合同研究集会で発表。

ⁱⁱ⁾ bcy10369@kwansei.ac.jp

〔表 1〕

	中国語の漢字列	返り点・堅点つき漢文	カッコ・堅点表示
①	国破山河在	国破山河在	国破山河在
②	烽火連三月	烽火連 _二 三月 _一	烽火連<三月>
③	感時花濺淚	感 _レ 時花濺 _レ 淚	感<時>花濺<淚>
④	借問新安吏	借 _一 _二 問新安吏 _一	借 _一 問<新安吏>
⑤	舉所佩玉玦	舉 _二 所 _レ 佩玉玦 _一	舉<所<佩>玉玦>
⑥	不以千里稱也	不 _下 以 _二 千里 _上 稱 _上 也	不<以<千里>稱>也
⑦	師不必賢於弟子	師不 _三 必賢 _二 於弟子 _一	師不<必賢<於弟子>>
⑧	加入寄乙位	加 _一 _三 入寄 _二 乙位 _一	加 _一 入<寄<乙位>>
⑨	渾欲不勝簪	渾欲 _レ 不 _レ 勝 _レ 簪	渾欲<不<勝<簪>>>
⑩	亡処亡氣	亡 _二 処亡 _レ 氣	亡<処亡<氣>>
⑪	有楚大夫於此	有 _レ 楚 _一 大 _一 _二 夫於此 _一	有<楚 _一 大 _一 夫<於此>>

以上から, $\mathbf{a} = a_1 a_2 \cdots a_n$ に対する, カッコ・堅点表示 $r_n(\mathbf{a}) = r_n(a_1 a_2 \cdots a_n)$ を次のように再帰的に定義するのが妥当であろう.

定義 A

$$r_n(\mathbf{a}) = \begin{cases} (1) & a_1 r_{n-1}(a_2 \cdots a_n) \\ (2) & a_1 < r_i(a_2 \cdots a_{i+1}) > r_{n-1-i}(a_{i+2} \cdots a_n) \\ (3) & a_1 \cdots - a_i < r_j(a_{i+1} \cdots a_{i+j}) > r_{n-i-j}(a_{i+j+1} \cdots a_n) \end{cases}$$

ただし, 中カッコ{ の(1)(2)(3)は「または」を意味し, (1)は $n \geq 1$, (2)(3)は $n \geq 2$ で成り立つとする. (1)は, $n = 1$ のとき $r_{n-1}(a_2 \cdots a_n)$ を empty つまり $r_1(a_1) = a_1$ とする. (2)は, $1 \leq i \leq n-1$ とし, $i = n-1$ のとき $r_{n-1-i}(a_{i+2} \cdots a_n)$ を empty とする. (3)は, $2 \leq i \leq n-1, 1 \leq j \leq n-i, i+j \leq n$ とし, $i+j = n$ のとき $r_{n-i-j}(a_{i+j+1} \cdots a_n)$ を empty とする.

2. 漢文数列の母関数

このように定義すると, (I)は次のように言い換えられる. (II)長さ n のストリング $\mathbf{a} = a_1 a_2 \cdots a_n$ に対する, カッコ・堅点表示 $r_n(\mathbf{a})$ の個数は幾つあるだろうか. 以下, n 次の r_n 全体の集合を $\mathbb{R}_n$, その元の個数を k_n とする. 場合の数を数えてみよう.

(1) $a_1 a_2$ 型るとき, a_1 は場合の数に無関係なので, 総数は k_{n-1} .

(2) $a_1 < a_2$ 型るとき, $\langle \rangle$ の中に i 字($1 \leq i \leq n-1$)あり, 残りは $n-1-i$ 字なので, 総数は $\sum_{i=1}^{n-1} k_i k_{n-1-i}$.

(3) $a_1 - a_2$ 型るとき, $-$ でつながれた漢字が i 字($2 \leq i \leq n-1$), $\langle \rangle$ の中に j 字($1 \leq j \leq n-i$)あり, 残りは $n-i-j$ 字なので, 総数は $\sum_{i=2}^{n-1} \sum_{j=1}^{n-i} k_j k_{n-i-j}$.

よって,

$$k_n = k_{n-1} + \sum_{i=1}^{n-1} k_i k_{n-1-i} + \sum_{i=2}^{n-1} \sum_{j=1}^{n-i} k_j k_{n-i-j} = k_{n-1} + \sum_{i=1}^{n-1} \sum_{j=1}^{n-i} k_j k_{n-i-j} .$$

上の式を漢文漸化式とよび, $\{k_n\}$ を漢文数列とよぶ.

定義 A において, 豎点を使わない(1)(2)だけに限定した場合の漢文漸化式は, (1)(2)の場合の数の和になる. k_n を c_n に, j を i に書きかえて,

$$c_n = c_{n-1} + \sum_{i=1}^{n-1} c_i c_{n-1-i} = \sum_{i=0}^{n-1} c_i c_{n-1-i} .$$

このように右辺がたたみ込み(convolution) $\sum_{i=0}^{n-1} c_i c_{n-1-i}$ だけになる場合, 一般項 c_n はカタラン数(Catalan number) $\frac{1}{n+1} \binom{2n}{n}$ になる([1])ⁱⁱⁱ⁾. つまりカタラン数は, 豎点を使うこともある場合の漢文数列の特殊ケースになっている. カタラン数の初項は $c_0 = 1$ とするので, 漢文数列においても $k_0 = 1$ とするのが妥当と言える.

なお, カタラン数の項比 c_{n+1}/c_n の極限は 4 になるので, 豎点を使わない場合, 「十分に大きな中国語の漢字列では, 漢字が 1 字増えても, 日本語への『翻訳可能性(translatability, interpretability)』は 4 倍程度に抑えられる」と主張できる.

次に, 漢文数列の母関数 $g(x) = \sum_{n=0}^{\infty} k_n x^n$ を求めよう.

漢文漸化式は, 次のように変形できる.

$$k_n = k_{n-1} + (k_1 k_{n-2} + \cdots + k_{n-1} k_0) + \cdots + (k_1 k_1 + k_2 k_0) + k_1 k_0 .$$

よって,

$$k_{n+1} = k_n + (k_1 k_{n-1} + \cdots + k_{n-1} k_1 + k_n k_0) + k_n - k_{n-1} .$$

これより, 次の漸化式を得る. ただし, $k_0 = k_1 = 1$ とする.

$$k_n = k_{n-1} - k_{n-2} + \sum_{i=0}^{n-1} k_i k_{n-1-i} .$$

この両辺に x^n を掛け, $n = 2$ から ∞ までの総和をとると, 母関数 $g(x)$ の 2 乗は $\sum_{n=0}^{\infty} (\sum_{i=0}^n k_i k_{n-i}) x^n$ だから, 次の 2 次の関数方程式を得る.

$$xg(x)^2 - (1 - x + x^2)g(x) + 1 - x = 0 .$$

よって,

定理 B

漢文数列 $\{k_n\}$ の母関数は次のようになる.

$$\begin{aligned} & \frac{1 - x + x^2 - \sqrt{1 - 6x + 7x^2 - 2x^3 + x^4}}{2x} \\ &= 1 + x + 2x^2 + 6x^3 + 20x^4 + 70x^5 + 254x^6 + 948x^7 + 3618x^8 + \cdots \end{aligned}$$

これは, M. D. Atkinson and T. Stitt[2]の[定理 17]で証明されている, 2 つの群 $\mathcal{U}$,

ⁱⁱⁱ⁾ 2009 年 9 月の計量国語学会で発表.

$\mathcal{S}$ のリース・プロダクト $\mathcal{U} \setminus \mathcal{S}$ の母関数と比較すると、初項を除いて一致している。
 本稿は、リース・プロダクトの概念を使わず、2 階の階差と、たたみ込み(convolution)
 $\sum_{i=0}^{n-1} k_i k_{n-1-i}$ が母関数の 2 乗の係数になることを利用している。
 なお、定義Aの $r_n(\mathbf{a})$ は、次のように一括して定義することもできる。

定義C

$$r_n(\mathbf{a}) = a_1 \cdots a_i < r_j(a_{i+1} \cdots a_{i+j}) > r_{n-i-j}(a_{i+j+1} \cdots a_n)$$

ただし、 $1 \leq i \leq n-1, 0 \leq j \leq n-1, i+j \leq n$ とする。

また、 $i=1$ のとき、(1) $j=0$ ならば、 $< r_j(a_{i+1} \cdots a_{i+j}) >$ を empty とし、(2) $j \geq 1$

かつ $i+j = n$ ならば、 $r_{n-i-j}(a_{i+j+1} \cdots a_n)$ を empty とする。

$i \geq 2$ のとき、(3) $i+j = n$ ならば、 $r_{n-i-j}(a_{i+j+1} \cdots a_n)$ を empty とする。

このように定義した $r_n(\mathbf{a})$ をカッコ・堅点表示の標準形とよぶ。

3. 「返読」がもたらす語順変換

前節までは、中国語の漢字列つまりストリング $\mathbf{a} = a_1 a_2 \cdots a_n$ にカッコの組と堅点を外部から付け加える、という観点から議論を進めてきた。ここからは、「返読」が生み出す中国語から日本語への語順変換を定義し、カッコ・堅点表示と 1 対 1 に対応していることを示そう。

まず、すべての「漢字(無定義用語)」の集合 $\Sigma = \{\sigma_1, \sigma_2, \dots, \sigma_n\}$ を考える。 Σ の元によるすべての順列を漢文言語 Σ^* とよび、その元を「漢字列」とよぶ。

$$\Sigma^* = \{\lambda, \sigma_1, \sigma_2, \dots, \sigma_n, \sigma_1 \sigma_1, \sigma_1 \sigma_2, \dots, \sigma_1 \sigma_n, \sigma_2 \sigma_1, \dots, \sigma_3 \sigma_1 \sigma_1 \sigma_3 \sigma_5, \dots\}$$

ここで、 λ は空順列を指し、「空の漢字列」とよぶ。

Σ^* の元 u, v を並べる 2 項演算 $u \cdot v$ を「連接(concatenation)」とよぶ。 $u \cdot v = uv$ 。

同じく、 u, v の順序を入れ替える 2 項演算 $u \langle v \rangle$ を「返接(英文名未定)」とよぶ。

$u \langle v \rangle = v \cdot u = vu$ 。ここで、 u が漢字 1 字ではなく、たとえば $u = \sigma_3 \sigma_1 \sigma_3 \sigma_2$ のとき、連結記号 $\cdot$ を用い、 $u \langle v \rangle$ を $\sigma_3 - \sigma_1 - \sigma_3 - \sigma_2 \langle v \rangle$ のように表記する。

以下、連接の $\cdot$ 記号を省略し、連接と返接が続くときは返接が優先するとする。

また、返接の演算をおこなうことを「カッコをはずす」とよぶ。上の例でカッコをはずすと、 $u \langle v \rangle = \sigma_3 - \sigma_1 - \sigma_3 - \sigma_2 \langle v \rangle = v \sigma_3 \sigma_1 \sigma_3 \sigma_2 = vu$ となる。

連接および返接では結合法則が成り立つほか、任意の $u \in \Sigma^*$ に対し、空の漢字列 λ が $\lambda u = u \lambda = u$ を満たすとする、 λ は連接と返接の単位元になる。ただし、連接、返接とも逆元は存在しない。すなわち漢文言語 Σ^* は、連接および返接の 2 つの 2 項演算に関して単位元をもった半群(semigroup)つまりモノイド(monoid)になっている。

なお、任意の $s, t, u, v, w \in \Sigma^*$ に対し、 $s \langle tu \langle v \rangle w \rangle$ のカッコを内側からはずすと、 $s \langle truw \rangle = truw s$ 、外側からはずしても、 $tu \langle v \rangle ws = truw s$ が成り立つので、 $s \langle tu \langle v \rangle w \rangle$ と書いても良いことがわかる。

以上にもとづき、定義A(またはそれと同等な定義C)のカッコ・堅点表示 $r_n(\mathbf{a})$ のカッコをすべてはずすと、 n 字の漢字列 $a_1 a_2 \cdots a_n$ を並べ替えた $j_n(\mathbf{a}) = a_{j_1} a_{j_2} \cdots a_{j_n}$ が得られる。 $j_n(\mathbf{a})$ は「返読」によって生み出される日本語の漢字列を意味している。

さて、(Ⅲ) n 字のカッコ・堅点表示を「返読」した、 $j_n(a)$ の総数は幾つあるだろうか。この問いの答えは、容易に求められる。なぜなら、もとの中国語の漢字列 $a_1 a_2 \cdots a_n$ と、「返読」後の日本語の漢字列 $j_n(a) = a_{j_1} a_{j_2} \cdots a_{j_n}$ が与えられたとき、次のような手順で $r_n(a)$ が復元できるからである。

1. $j_n(a)$ の中で a_1 を探す。
2. a_1 が左端($a_{j_1} = a_1$)にあれば、 $r_n(a) = a_1 a_2 \cdots$ とする。
3. a_1 が左端になければ、 a_2 を探し、
 - ① a_2 が a_1 の右隣にあれば、 $r_n(a) = a_1 - a_2 \cdots$ とする。
 - ② a_2 が a_1 の左側にあり、 a_m を最大として連続した添数をもつ漢字がすべて a_1 の左側にあれば、 $r_n(a) = a_1 < a_2 \cdots a_m > \cdots$ とする。
 - ③ その他の場合、 $j_n(a)$ になる a のカッコ・堅点表示は存在しない。
4. $j_n(a)$ から a_1 を取り除き、 a_2 を探す。・・・といった手順を繰り返す。

すなわち、与えられた $j_n(a)$ から $r_n(a)$ は一意的に決定できる。よって、

定理 D

$r_n(a)$ と $j_n(a)$ は 1 対 1 に対応している。

これより、 $j_n(a)$ の総数は $r_n(a)$ の総数すなわち漢文数列 $\{k_n\}$ と等しく、その母関数は定理 B で与えられることがわかる。

$j_n(a)$ は、カッコの組と堅点つまり返接演算で制限された語順変換であるので、一定の制限をつけた順列や順列行列、グラフ(木)などで定義できる([3])。

たとえば、 n 次の順列行列全体の集合を $\mathbb{P}_n$ とし、 $A \in \mathbb{P}_n, B \in \mathbb{P}_m$ のとき、直和 $A \oplus B$ と直差 $A \ominus B$ を下のように定義すると、 $A \oplus B, A \ominus B \in \mathbb{P}_{n+m}$ となる。

$$A \oplus B = \begin{pmatrix} A & \\ & B \end{pmatrix}, A \ominus B = \begin{pmatrix} & A \\ B & \end{pmatrix}.$$

さらに、順列行列全体の集合を $\mathbb{P}$ 、 n 次の単位行列を E_n として、 $\mathbb{P}$ の部分集合である漢文行列の集合 $\mathbb{H}$ を次の(1)(2)(3)で再帰的に定義する。なお、ここでは、中国語の漢字列を行ベクトルとみなし、その右から漢文行列を掛けることにしている。

定義 E

- (1) $E_1 \in \mathbb{H}$
- (2) $A \in \mathbb{H}$ ならば、 $E_1 \oplus A \in \mathbb{H}$
- (3) $A, B \in \mathbb{H}$ ならば、任意の n に対し、 $E_n \ominus A \in \mathbb{H}$ かつ $(E_n \ominus A) \oplus B \in \mathbb{H}$

このように定義した漢文行列の集合 $\mathbb{H}$ のうち、 n 次の漢文行列の集合を $\mathbb{H}_n$ とすると、 $\mathbb{H}_n$ の元の個数は k_n に等しく、 n 次のカッコ・堅点表示全体の集合 $\mathbb{R}_n$ と同型になる。むしろ $\mathbb{H}$ は $\mathbb{R} = \bigcup_{n \in \mathbb{N}} \mathbb{R}_n$ と同型になる。

定理 F

$\mathbb{H}_n \simeq \mathbb{R}_n$ かつ $\mathbb{H} \simeq \mathbb{R}$

現在, $\mathbb{H}$ の, 直和および直差に関する代数的構造の一部は明らかになっており, とくに $\mathbb{H}_n$ の元 (漢文行列) のブロック対角化および巡回群^{iv)} について一定の結果を得ているが, 発表の段階に至っていない.

4. 今後の課題

本稿の核心は, 中日 2 言語間に横たわる返り点・堅点システムを, 入れ子構造をもつ「カッコ・堅点表示」として形式化・抽象化した点にある.

当面の課題としては, ここで母関数を得た漢文数列 $\{k_n\}$ の一般項や, 「翻訳可能性」に相当する項比の極限などを算出しなければならない.

また, 返接 $\beta < \alpha >$ における β の堅点の数を 0 個 (単返接モデルとよぶ), 1 個, 2 個, $\dots$, m 個までに制限した, m 次の漢文数列 $\{k_{m,n}\}$ の母関数は容易に求められるので, 返り点や堅点の使用頻度の調査をおこない, $\{k_{m,n}\}$ と照合して, 統計的に分析する作業が残っている. 実際の漢文訓読の文例では, 堅点を使うケースは少なく, ほとんどが単返接モデルになっている.

なお, 中英や英日の翻訳では, 中日の間すなわち漢文訓読での堅点のような制限が返接になく, 返って読まれる列のなかでも返接が存在する. このような 2 言語間の翻訳における語順変換をどう捉えるかは今後の大きな課題となる.

より一般的には, $\mathbb{R}$ や $\mathbb{H}$ をはじめとする“多次元”のメタ言語空間の「位相」を研究し, 「なぜ返読という操作が生まれたのか」を理論的に明らかにする必要があると思う.

参考文献

- [1] R. P. Stanley, Enumerative Combinatorics, Vol. I, II, Cambridge University Press, 1999
- [2] M. D. Atkinson and T. Stitt, Restricted permutations and the wreath product, Discrete Math., 259 (2002), 19–36.
- [3] D. E. Knuth, The Art of Computer Programming, Vol. I / Fundamental Algorithms 3rd ed., Addison–Wesley, 1997

^{iv)} 任意の n 次の順列行列 A は適当な n 次の順列行列 P を選び $P^T A P = J_{r_1} \oplus J_{r_2} \oplus \dots \oplus J_{r_i}$ とできる. ここで, J_r は r 次の巡回行列 ($J_r = E_1 \ominus E_{r-1}$) で, $1 \leq r_1 \leq r_2 \leq \dots \leq r_i \leq n$ かつ $r_1 + r_2 + \dots + r_i = n$. このとき A を生成元とする巡回群の位数は $r_1, r_2, \dots, r_i$ の最小公倍数になる. これより, たとえば $7(=3+4)$ 次の順列行列には 12 乗しなければ単位行列にならないものが存在することがわかる. つまり「自然数 n を幾つかの自然数に分割し, それらの最小公倍数のうち最大のもの (最大の最小公倍数) を求める問題」がここから派生する.