2012年　計量国語学会　予稿

漢文訓読における返り点・竪(たて)点システムの数学的構造について

島野達雄

1.はじめに

漢文訓読とは，中国語の漢字列に返り点，竪点などを付け加えて日本語化する「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，返り点や竪点などの付け方も一様ではない．返り点や竪点は，「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と言えるであろう．

本稿では，2009年の本学会での発表に引き続き，返り点と竪点がもたらす「返読」に焦点をあて，漢文訓読の基本構造を理論的に明らかにしたい．

前半は，いくつかの例文をあげ，(1)返り点・竪点つき漢文が，もとの中国語の漢字列にカッコの組（＜＞）と竪点（－）を付け加えた，「カッコ・竪点表示」に書きかえられること， (2)このカッコの組と竪点を順にはずしてゆき，日本語の漢字列に並べかえる操作が「返読」に相当する語順変換であること，の2点を中心に説明する．

後半は，「カッコ・竪点表示」を数学的に定義し，(2)のカッコ・竪点表示の総数すなわち語順変換の総数に関する漸化式を導き，2009年に発表した竪点を使わないケースを含む，竪点を使うこともある一般的な場合の母関数を証明するほか，今後の課題を提起する．

2.返り点・竪点の働き

はじめに，説明の便宜上，「β＜α＞」と「カッコ・竪点表示」を次のように規定する．

返り点や竪点が付いている漢文において，βが「返って読む，漢字1字または竪点でつながれた2字以上の漢字」，αが「βより先に読む部分」であれば，β＜α＞と書くこととし，「αとβの返接」とよぶ．返り点と竪点を返接β＜α＞の形式に置きかえたものを，返り点・竪点つき漢文のカッコ・竪点表示とよぶ．

〔表 1〕の①「国破山河在」のように，返り点・竪点がまったく付いていない漢文は，そのままの「国破山河在」をカッコ・竪点表示とよぶことにする．

②「烽火連$_{二}$三月$_{一}$」では，「烽火」はそのままで，「連」は返って読むβに，「三月」はβより先に読むαに相当するので，カッコ・竪点表示は「烽火連＜三月＞」となる．日本語の文法では，βの「連」が動詞，αの「三月」が補語と言えよう．

③「感$_{レ}$時花濺$_{レ}$涙」はレ点がある場合で，「感＜時＞花濺＜涙＞」となる．2つのレ点は，βの「感」「濺」が動詞，αの「時」が補語，「涙」が目的語になっている．⑧の前段の「倍$_{レ}$之」も「倍＜之＞」と，βの「倍」が動詞，αの「之」が目的語である．

④「借－$_{二}$問新安吏$_{一}$」では返って読むβが竪点でつながれており，「借－問＜新安吏＞」となる．βの「借－問」が動詞，αの「新安吏」が補語になっている．

〔表 1〕

	返り点・竪点つき漢文	カッコ・竪点表示	日本語の漢字列	順列
①	国破山河在	国破山河在	国破山河在	12345
②	烽火連㆓三月㆒	烽火連＜三月＞	烽火三月連	12534
③	感㆑時花濺㆑涙	感＜時＞花濺＜涙＞	時感花涙濺	21354
④	借－㆓問新安吏㆒	借－問＜新安吏＞	新安吏借問	45123
⑤	挙㆓所㆑佩玉玦㆒	挙＜所＜佩＞玉玦＞	佩所玉玦挙	52134
⑥	不㆘以㆓千里㆒称㆖也	不＜以＜千里＞称＞也	千里以称不也	531246
⑦	師不㆔必賢㆓於弟子㆒	師不＜必賢＜於弟子＞＞	師必於弟子賢不	1726345
⑧	倍㆑之加－㆔入寄㆓乙位㆒	倍＜之＞加－入＜寄＜乙位＞＞	之倍乙位寄加入	2167534
⑨	渾欲㆑不㆑勝㆑簪	渾欲＜不＜勝＜簪＞＞＞	渾簪勝不欲	15432
⑩	亡㆓処亡㆒㆑気	亡＜処亡＜気＞＞	処気亡亡	4132
⑪	有㆑楚－大－㆓夫於此㆒	有＜楚－大－夫＜於此＞＞	於此楚大夫有	634512

⑤は一二点の中にレ点がある場合．「所㆑佩」のカッコ・竪点表示は「所＜佩＞」で，意味の上では，αの「佩びること」を対象にして，βの「するところ」が作用している，と見なすことができる．この「所㆑佩」で「玉玦」を修飾した「所＜佩＞玉玦」が＜＞でくくられ，βの動詞「挙」の目的語となっている．

⑥は上下点の中に一二点がある場合．上と同様に，βの「以」が動詞，αの「千里」が目的語となって，「以＜千里＞」となり，さらに，βの「不」がαの「以＜千里＞称」を対象として否定している．「也」は返り点が付いていないので，そのままにしている．

⑤⑥は単純な入れ子構造をもった文と言えるであろう．

⑦の「不㆔必賢㆓於弟子㆒」の「不」は「必賢㆓於弟子㆒」を対象語として否定している，と見なすことができる．「賢㆓於弟子㆒」のカッコ・竪点表示は「賢＜於弟子＞」であるので，全体として「不㆔必賢㆓於弟子㆒」は「不」がβ，「必賢＜於弟子＞」がαとなり，カッコ・竪点表示は「師不＜必賢＜於弟子＞＞」と，2 組の＜＞が入れ子になる．

⑧の後段の「加－㆔入寄㆓乙位㆒」も同様で，動詞「加－入」がβ，「寄㆓乙位㆒」すなわち「寄＜乙位＞」がαと見なせる．興味深いのは，「寄＜乙位＞」はβの「寄」が動詞，αの「乙位」が補語だが，「寄＜乙位＞」がβの「加－入」のカッコ内つまりαになるときは，「乙位に寄せるものを」と目的語化することである．

⑨のレ点の連続でも同様に，「欲」の目的語が「不㆑勝㆑簪」，「不」が否定する目的語が「勝㆑簪」，「勝」の目的語が「簪」で，カッコ・竪点表示は「欲＜不＜勝＜簪＞＞＞」と三重の入れ子になる．

⑩「亡㆓処亡㆒㆑気」は，一レ点とよばれることもある一とレ点の複合，⑪「有㆑楚－大

－$_{\text{二}}$夫於此$_{\text{一}}$」はレ点・一二点・竪点の複合だが，「亡$_3$処亡$_2$気$_1$」「有$_3$楚－大－夫$_2$於此$_1$」とアラビア数字の 123 点に書きかえることができるので，⑦と同様に，「亡＜処亡＜気＞＞」「有＜楚－大－夫＜於此＞＞」とカッコ・竪点表示にできる．

以上，おもな返り点・竪点つき漢文がカッコ・竪点表示に書きかえられることを示した．

カッコの組（＜＞）は，一二点の中のレ点などのほか，一二三点やレ点の連続などでも入れ子になり，どのβ＜α＞においても，言い方はともかく，βは作用語（動詞，「不」など），αは被作用語（対象語，目的語，補語など）になっている．

3.返読の基本構造

次に，カッコ・竪点表示にふくまれているβ＜α＞のカッコをはずした後，βの竪点を取り除く操作が，「返読」に相当する語順変換（漢字の並べ替え）であることを示そう．

αとβの返接β＜α＞のカッコをはずす，とは，β＜α＞をαβに置きかえ，その後，βに竪点がある場合はそれを取り除くことをいう．以下，βとβの竪点を取り除いたものを同一視し，β＜α＞＝αβと書く．

なお，γ＜β＜α＞＞のカッコを内側からはずすとγ＜αβ＞＝αβγ，外側からはずしてもβ＜α＞γ＝αβγなので，カッコは内側，外側のどちらから先にはずしてもよい．

たとえば，〔表 1〕の⑤「挙＜所＜佩＞玉玦＞」は内側のカッコから「挙＜佩所玉玦＞」とはずし，その後，外側のカッコをはずすと「佩所玉玦挙」となるが，外側を先にはずして「所＜佩＞玉玦挙」，次に内側をはずしても「佩所玉玦挙」と同じ結果になる．

さて，n字の中国語の漢字列$\boldsymbol{a} = a_1 a_2 \cdots a_n$にカッコと竪点を付け加えたカッコ・竪点表示のすべてのカッコをはずすと，最終的に$\boldsymbol{a}$を並べ替えた漢字列$j(\boldsymbol{a}) = a_{j_1} a_{j_2} \cdots a_{j_n}$を得る．

〔表 1〕の「日本語の漢字列」は，この$j(\boldsymbol{a}) = a_{j_1} a_{j_2} \cdots a_{j_n}$を示している．言うまでもなく，この漢字列は「返り点・竪点つき漢文」を「返読」した結果でもある．

すなわち，カッコ・竪点表示は，n 次元のベクトル$\boldsymbol{a} = a_1 a_2 \cdots a_n$を変数とする$j\colon \boldsymbol{a} \to j(\boldsymbol{a})$という語順変換（漢字の並べ替え）をおこなう関数を意味しており，順列行列や順列で表現ができる．たとえば，「表－章＜之＞＝之表章」は，

(表章之)$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$=(之表章)から，順列行列$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$であらわせ，順列は 2 3 1 になる．

2009 年の発表では，カッコ・竪点表示の数学的な定義とあわせて，順列行列とグラフによる，「訓読行列」「訓読の木」の定義を紹介したが，ここでは省略する．

ちなみに，異なるn字の変数$\boldsymbol{a}$が与えられた場合，$j(\boldsymbol{a})$からカッコ・竪点表示が復元できる．最初の 2 字の関係は，$a_1 a_2$，$a_1 < a_2$，$a_1 – a_2$の 3 通りだけなので，「$j(\boldsymbol{a})$の中で最初にa_1を探し，左端にあれば$a_1 a_2$とする．a_1が左端になければa_2を探し，a_1の左側にあれば$a_1 < a_2$とし，a_1の右隣にあれば$a_1 - a_2$とする」といった手順をa_1から順に繰り返せばよい．

⑩「亡＜処亡＜気＞＞＝処気亡亡」のように同じ漢字が変数$\boldsymbol{a}$にある場合は，上の手順ではカッコ・竪点表示を復元できない．

また，カッコ・竪点表示の語順変換 $j: \boldsymbol{a} \to j(\boldsymbol{a})$ の逆変換 j^{-1} は必ず存在するが，必ずしもカッコ・竪点表示の語順変換にはならないことを注意しておく．

4.訓読漸化式とカタラン数

「カッコ・竪点表示」は，次のように再帰的に定義できる．

漢字列$\boldsymbol{a} = a_1 a_2 \cdots a_n$に，「左カッコ（<）の左側は漢字 1 字または竪点（－）でつながれた 2 字以上の漢字」というルールで，0 個以上のカッコの組（<>）と竪点（－）を付け加える操作をあらわす関数$r_n: \boldsymbol{a} \to \boldsymbol{r_n(a)}$を考える．

ここで，すべてのr_nの集合を$\mathbb{R}_n$と書き，「n字（n次）のカッコ・竪点表示」とよぶ．

$n = 1$のとき，$\boldsymbol{a} = a_1$にカッコや竪点は付けられず，$\mathbb{R}_1$には$r_1(\boldsymbol{a}) = a_1$の 1 つしかない．

$n \geq 2$のとき，$r_n(\boldsymbol{a}) = r_n(a_1 a_2 \cdots a_n)$の最初の 2 字の関係は，$a_1 a_2$，$a_1 < a_2$，$a_1 - a_2$の 3 通りだけなので，次の❶❷❸のように場合分けして，r_nを再帰的に定義する．

❶$a_1 a_2$のとき，$r_n(a_1 a_2 \cdots a_n) = a_1 r_{n-1}(a_2 \cdots a_n)$とする．

❷$a_1 < a_2$のとき，$r_n(a_1 a_2 \cdots a_n) = a_1 < r_i(a_2 \cdots a_{i+1}) > r_{n-1-i}(a_{i+2} \cdots a_n)$とする．
ただし，$1 \leq i \leq n-1$で，$i = n-1$のとき$r_{n-1-i}(a_{i+2} \cdots a_n)$は空（empty）とする．

❸$a_1 - a_2$のとき，$r_n(a_1 a_2 \cdots a_n) = a_1 \cdots - a_i < r_j\left(a_{i+1} \cdots a_{i+j}\right) > r_{n-i-j}(a_{i+j+1} \cdots a_n)$とする．
ただし，$2 \leq i \leq n-1$，$1 \leq j \leq n-i$で，$i + j = n$のとき$r_{n-i-j}(a_{i+j+1} \cdots a_n)$は空(empty)とする．

この定義にもとづき，$\mathbb{R}_n$の元の個数k_nを数え上げてみよう．便宜上，$k_0 = 1$とする．

⑴$a_1 a_2$のとき，a_1は場合の数に無関係なので，総数はk_{n-1}．

⑵$a_1 < a_2$のとき，<>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$\sum_{i=1}^{n-1} k_i k_{n-1-i}$．

⑶$a_1 - a_2$のとき，－でつながれた漢字がi字$(2 \leq i \leq n-1)$，<>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，

$$k_n = k_{n-1} + \sum_{i=1}^{n-1} k_i k_{n-1-i} + \sum_{i=2}^{n-1} \sum_{j=1}^{n-i} k_j k_{n-i-j} = k_{n-1} + \sum_{i=1}^{n-1} \sum_{j=1}^{n-i} k_j k_{n-i-j} \ . \qquad (4.1)$$

上の式を訓読漸化式とよび，$\{k_n\}$を漢文数列とよぶ．k_nは「n字の漢字列にカッコの組と竪点を付け加えるパターンが何通りあるか」を示している，と言えよう．

また，以上のように返接β＜α＞のβに竪点を使うこともある場合を「連返接モデル」とよび，竪点をまったく使わない場合を「単返接モデル」とよぶ．

単返接モデルの場合の訓読漸化式は，先の⑴⑵の和になる．k_nをc_nに書きかえて，

$$c_n = c_{n-1} + \sum_{i=1}^{n-1} c_i c_{n-1-i} = \sum_{i=0}^{n-1} c_i c_{n-1-i}\ . \tag{4.2}$$

2009 年の発表では，こののc_nの母関数 $\sum_{n=0}^{\infty} c_n x^n$ を算出し，一般項がカタラン数$\frac{1}{n+1}\binom{2n}{n}$になることを証明した．

以上のことから，漢文数列がカタラン数の拡張になっていることがわかる．一般に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．

5.漢文数列の母関数

(4.1)の訓読漸化式は，次のように変形できる．

$$\begin{aligned} k_n &= k_{n-1} + \sum_{j=1}^{n-1} k_j k_{n-1-j} + \cdots + \sum_{j=1}^{2} k_j k_{2-j} + \sum_{j=1}^{1} k_j k_{1-j} \\ &= k_{n-1} + (k_1 k_{n-2} + \cdots + k_{n-2}k_1 + k_{n-1}k_0) + \cdots + (k_1 k_1 + k_2 k_0) + k_1 k_0. \end{aligned} \tag{5.1}$$

よって，

$$\begin{aligned} k_{n+1} &= k_n + (k_1 k_{n-1} + \cdots + k_{n-1}k_1 + k_n k_0) + (k_1 k_{n-2} + \cdots + k_{n-2}k_1 + k_{n-1}k_0) + \cdots + k_1 k_0 \\ &= k_n + (k_1 k_{n-1} + \cdots + k_{n-1}k_1 + k_n k_0) + k_n - k_{n-1} \\ &= \sum_{i=0}^{n} k_i k_{n-i} + k_n - k_{n-1}. \end{aligned} \tag{5.2}$$

これより，次の漸化式を得る．ただし，$k_0 = k_1 = 1$とする．

$$k_n = k_{n-1} - k_{n-2} + \sum_{i=0}^{n-1} k_i k_{n-1-i}\ . \tag{5.3}$$

この両辺に x^n を掛け，$\mathrm{n} = 2$から∞までの総和をとると，母関数 $g(x) = \sum_{n=0}^{\infty} k_n x^n$ 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．

$$xg(x)^2 - (1 - x + x^2)g(x) + 1 - x = 0\ . \tag{5.4}$$

2 次方程式の解の公式で得られる複号の正のものは，初項が定数項でないので不適．

よって，漢文数列 $\{k_n\}$ の母関数は次のようになる．

$$\begin{aligned} &\frac{1 - x + x^2 - \sqrt{1 - 6x + 7x^2 - 2x^3 + x^4}}{2x} \\ &= 1 + x + 2x^2 + 6x^3 + 20x^4 + 70x^5 + 254x^6 + 948x^7 + 3618x^8 + 14058x^9 \cdots. \end{aligned} \tag{5.5}$$

これは，“M. D. Atkinson and T. Stitt; Restricted permutations and the wreath product, in Discrete Mathematics 259(2002)”の〔定理 17〕で証明されている，2 つの群 $\mathcal{U}$，$\mathcal{I}$ のリー

ス・プロダクト $\mathcal{U} \wr \mathcal{I}$ の母関数と比較すると，初項を除いて一致している．この論文を教示いただいた D. G. Rogers 氏に感謝する．

本稿の証明は，リース・プロダクトの概念を使わず，2 階の階差，および，$\sum_{i=0}^{n-1} k_i k_{n-1-i}$ が母関数の 2 乗の係数になることを利用している．

6.今後の課題

最後に，関連する今後の課題を提起しておきたい．

本稿の核心は，中国語と日本語という 2 つの言語の間に横たわる，返り点・竪点システムを，入れ子構造をもつ「カッコ・竪点表示」として形式化・抽象化した点にある．

カッコ・竪点表示は，漢字列$\boldsymbol{a} = a_1 a_2 \cdots a_n$に対する，$j: \boldsymbol{a} \to j(\boldsymbol{a})$という語順変換と見なせるほか，直接，0 個以上のカッコの組と竪点を$\boldsymbol{a}$に付け加える，$r_n: \boldsymbol{a} \to r_n(\boldsymbol{a})$として再帰的に定義することもできる．本稿では，この再帰的定義にもとづき，すべてのr_nの集合$\mathbb{R}_n$の元の個数k_nを数え上げている．

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，一般項や項比の極限などを算出しなければならないだろう．

また，本稿では，返接 β＜α＞の β が「漢字 1 字または竪点でつながれた 2 字以上の漢字」という制限もつカッコ・竪点表示を「連返接モデル」とよび，より厳しく漢字 1 字に限定したカッコ・竪点表示を「単返接モデル」とよんでいる．

この返接の β の竪点の数を 0 個（＝単返接モデル）,1 個,⋯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，第二の当面の課題としては，返り点や竪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ある．

今回の議論をさらに進めれば，中国語の漢字列に返り点・竪点を付け加える，という操作の全体は，$\mathbb{R}_n$の合併集合$\bigcup_{n=1}^{\infty} \mathbb{R}_n$であらわされると言える．

今後，このような“多次元”の関数空間の距離，言うなれば異なった次元をもつ語順変換からなる空間の「位相」を研究し，漢文訓読の本質を明らかにする必要があると思う．

なお，2009 年の発表でもふれたように，中英や英日の翻訳では，中日の間すなわち漢文訓読での竪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，人工言語をふくむメタ言語の「位相」を定式化し，本質を明らかにすることが，これからの研究目標になるであろう．