2022 年 9 月 17 日（土）計量国語学会（対面／オンライン）
於　日本大学文理学部

漢文訓読における交点数（返読量）の公式

島野達雄（関西学院大学）[1]

2019 年の本学会では，二つのn次順列の「距離」を交点の数で定義して可視化し，順列αと単位順列εとの距離を「順列αの交点数」と定義した[2]．漢文訓読においては，語順変換に対応する順列（K 順列）の交点数は，返読量（返読する延べ漢字数）を意味している[3]．

本発表では，一般の順列の交点数とあわせて，漢文訓読の場合の交点数（返読量）の総和と平均を算出する．これらは，AI による古典籍の自動訓読システムの開発にあたって，実際の漢文の特徴量としての交点数（返読量）を分析する基準値となる．

１．n 次順列の交点数

n次順列$\alpha = \alpha_1\alpha_2\cdots\alpha_{n-1}\alpha_n$が与えられたとき，逆順の$\alpha^G = \alpha_n\alpha_{n-1}\cdots\alpha_2\alpha_1$ がただ一つ決まる．αとα^Gの交点数$d(\alpha,\alpha^G)$は，交点数の性質から次の式を満たす[4]．

$$d(\alpha,\alpha^G) = \frac{n(n-1)}{2} \qquad \cdots(1.1)$$

$n \geq 2$ のとき(α,α^G)のペアは$\dfrac{n!}{2}$組ある．よってn次順列の交点数の総和S_nと平均μ_nは，

$$S_n = \frac{n(n-1)}{2} \times \frac{n!}{2} = \frac{n(n-1)n!}{4} \qquad \cdots(1.2)$$

$$\mu_n = \frac{n(n-1)n!}{4} \div n! = \frac{n(n-1)}{4} \qquad \cdots(1.3)$$

となる[5]．

２．漢文訓読の数学モデル

はじめに，漢文訓読に用いられるレ点，一二点，上下点などの返り点と竪点（たててん．ハイフン，連続符号）を，左右一対の括弧$<\ >$とハイフン$-$に抽象化した数学モデル「括弧表示」の定義を述べる．

n個の漢字（無定義用語）を並べた列$\boldsymbol{a} = A_1A_2\cdots A_n$に対し，$n$字括弧表示$r_n(\boldsymbol{a})$を次のように再帰的に定義する[6]．empty は空，何も無いの意．0 字括弧表示は empty とする．

$$r_n(\boldsymbol{a}) = A_1\cdots - A_i < r_j(A_{i+1}\cdots A_{i+j}) > r_{n-i-j}(A_{i+j+1}\cdots A_n) \qquad \cdots(2.1)$$

ただし，$1 \leq i \leq n-1$，$0 \leq j \leq n-1$，$i+j \leq n$とする．$i=1$のとき，$A_1\cdots - A_i$はA_1とする．

$r_1(A_1) = A_1$とする．つまり漢字 1 字は括弧表示とみなす．

$i = 1$ かつ$j = 0$ならば，$< r_j(A_{i+1}\cdots A_{i+j}) >$を empty とする．

$i \geq 2$ のとき，$j \neq 0$ とする．かつ$i+j = n$ならば，$r_{n-i-j}\left(A_{i+j+1}\cdots A_n\right)$を empty とする．

$A_1 - A_2 \cdots - A_i$をi字連結項$(1 \leq i \leq n-1)$または単に連結項とよぶ．左括弧$<$の左側には，必ずi字連結項が存在する．$i = 1$のときの連結項はハイフンを用いないA_1になる．

再帰的とは，$r_n(\boldsymbol{a})$の定義のなかに$r_j\left(A_{i+1} \cdots A_{i+j}\right)$や$r_{n-i-j}\left(A_{i+j+1} \cdots A_n\right)$があらわれることをいう．たとえば$n = 8, i = 2, j = 4$のとき$r_8(A_1 A_2 \cdots A_8) = A_1 - A_2 < r_4(A_3 \cdots A_6) > r_2(A_7 A_8)$において$r_4(A_3 \cdots A_6) = A_3 - A_4 - A_5 < A_6 >$，$r_2(A_7 A_8) = A_7 < A_8 >$とすると，

$$r_8(A_1 A_2 \cdots A_8) = A_1 - A_2 < A_3 - A_4 - A_5 < A_6 \gg A_7 < A_8 > \qquad \cdots (2.2)$$

のように，$<\ >$の中に$<\ >$があらわれ，右括弧$>$に続いて（並置して），$A_7 < A_8 >$という括弧表示があらわれることをいう．

３．Ｋ順列の定義

n字括弧表示は次の手順でn次順列とみなすことができる．

⑴括弧表示の各漢字を各漢字の添え字に書き換える．(2.1)は，次のようになる．

$$r_n(\boldsymbol{a}) = 1 - \cdots - i < r_j(i+1, \cdots, i+j) > r_{n-i-j}(i+j+1, \cdots, n) \qquad \cdots (3.1)$$

⑵$<\ >$内の数字を$-$でつながれた数字列の前（左側）に並べる．この操作を「($<\ >$内を）返読する」，「括弧をはずす」，「括弧表示を展開する」などとよぶ．

⑶最後に$-$でつながれた数字列の$-$を取り除く．

たとえば，(2.2)は次のようにして順列とみなすことができる．

$$\begin{aligned} r_8(12 \cdots 8) &= 1 - 2 < 3 - 4 - 5 < 6 \gg 7 < 8 > \\ &= 1 - 2 < 63 - 4 - 5 > 87 \\ &= 63451287 \qquad \cdots (3.2) \end{aligned}$$

ここでは，二重括弧の内側の$3 - 4 - 5 < 6 >$を先に展開し，外側の$1 - 2 < 63 - 4 - 5\ >$を後で展開しているが，外側の括弧を先にはずし，内側の括弧を後からはずしても同じ結果になる．括弧の展開をおこなう二項演算は結合法則が成り立つ[7].

n字括弧表示を展開して得られる順列をn次 K 順列とよぶ．K順列から括弧表示は容易に復元できるので，括弧表示とK順列は 1 対 1 に対応している．括弧表示の返読量（$<\ >$内の延べ漢字数）は K 順列（およびK順列の逆順列）の交点数に等しい[8].

n次 K 順列 σ の交点数 $d(\varepsilon, \sigma)$の最小値は，σが単位順列εのときの$d(\varepsilon, \varepsilon) = 0$．このとき，括弧表示は$<\ >$と$-$をまったく用いない，

$$r_n(A_1 A_2 \cdots A_n) = A_1 A_2 \cdots A_n \qquad \cdots (3.3)$$ である．

n 次 K 順列の交点数の最大値は，εの逆順ε^Gの交点数$d(\varepsilon, \varepsilon^G) = n(n-1)/2$．括弧表示は，

$$r_n(A_1 A_2 \cdots A_n) = A_1 < A_2 < \cdots < A_n > \cdots >> \qquad \cdots (3.4)$$ である．

n次順列でも，交点数の最小値は$d(\varepsilon, \varepsilon) = 0$，最大値は$d(\varepsilon, \varepsilon^G) = n(n-1)/2$ になる．

なお，4 次順列のうち 2413，3142，2431，3241 は，$<\ >$と$-$をどのように組み合せても括弧表示の展開からは生じない．つまり，K 順列は順列全体の集合の真部分集合である．

４．括弧表示のパターン数

次にn字括弧表示のパターン数，すなわちＫ順列が幾つあるかを数え上げてみよう．以下，漢字列を AB…のようにアルファベット大文字であらわす．

初めにハイフンを用いない（連結項のハイフンの個数 0 個の）n字括弧表示のパターン数を$k_0(n)$とする．$-$を用いないので，括弧表示の先頭の 2 漢字の関係は，AB か A<B のどちらかである。AB 型の場合，パターン数は B 以下の$n-1$字のパターン数$k_0(n-1)$だけある．A<B 型の場合，< >内にj字$(1 \le j \le n-1)$があるとすると，>の右側に残りの$n-j-1$字があるので，パターン数は$\sum_{j=1}^{n-1} k_0(j)k_0(n-j-1)$になる．

よって，$k_0(0)=1$として，

$$k_0(n) = k_0(n-1) + \sum_{j=1}^{n-1} k_0(j)k_0(n-j-1) = \sum_{j=0}^{n-1} k_0(j)k_0(n-j-1) \qquad \cdots (4.1).$$

$\sum_{j=0}^{n-1} k_0(j)k_0(n-j-1)$を畳み込み（convolution）とよぶ．数列$k_0(n)$はカタラン数$\frac{1}{n+1}\binom{2n}{n}$であることが母関数$g_0(x)=\sum_{n=o}^{\infty} k_0(n)x^n$を利用して証明できる[9].

定義(2.1)で連結項$A_1 - A_2 \cdots - A_i$のハイフンの個数をm以下$(0 \le m \le n-2)$に制限すると，連結項の字数iは$m+1$以下に制限される．このとき，括弧表示のパターン数$k_m(n)$も，上に述べた$m=0$のカタラン数$k_0(n)$と同様に，先頭の 2 漢字の型によって求められる．すなわち，AB 型，A<B 型に加えて A$-$B 型に場合分けをすればよい．

AB 型のパターン数は$k_m(n-1)$.

A<B 型のパターン数は$\sum_{j=1}^{n-1} k_m(j)k_m(n-j-1)$.

A$-$B 型のとき，$-$でつながれた漢字がi字$(2 \le i \le m+1)$， < >の中にj字$(1 \le j \le n-i)$あるとすると，残りは$n-i-j$字なので，総数は，$\sum_{i=2}^{m+1}\sum_{j=1}^{n-i} k_m(j)\,k_m(n-i-j)$.

よって，

$$k_m(n) = k_m(n-1) + \sum_{j=1}^{n-1} k_m(j)k_m(n-j-1) + \sum_{i=2}^{m+1}\sum_{j=1}^{n-i} k_m(j)k_m(n-i-j) \qquad \cdots (4.2).$$

母関数$g_m(x)=\sum_{n=0}^{\infty} k_m(n)x^n$は、次の関数 2 次方程式の解になる．

$$x(1+x+x^2+\cdots+x^m){g_m}^2 - (1+x^2+x^3+\cdots+x^{m+1})g_m + 1 = 0 \qquad \cdots (4.3).$$

なお，nが与えられたとき，連結項のハイフンは高々$n-2$個しかありえない．nごとに決まる，連結項のハイフンが$n-2$個以下の括弧表示のパターン数$k_{n-2}(n)$をとくに$k(n)$と書く（$k(0)=k(1)=1$とする）．

$k(n)$の母関数$g(x)=\sum_{n=0}^{\infty} k(n)x^n$の関数 2 次方程式は，$xg^2-(1-x+x^2)g+1-x=0$となる．これは(4.3)の$m \to \infty$とした極限となっている[10].

５．交点数（返読量）の総和の公式

ハイフンを用いないn字括弧表示の返読量$H_0(n)$は，下の式となる（$H_0(0)=0$とする）．

$$H_0(n)=H_0(n-1)+\sum_{j=1}^{n-1}\{jk_0(j)k_0(n-j-1)+H_0(j)k_0(n-j-1)+H_0(n-j-1)k_0(j)\} \quad \cdots(5.1).$$

右辺の$H_0(n-1)$は，先頭の 2 漢字が AB 型のとき，交点数の和は，B 以下の$n-1$字の括弧表示の交点数の和に等しいことを示している．

A<B 型のときは，A< B⋯ >の< >のなかに j 字（$1\le j\le n-1$）があるとすると，Σの初めにある$jk_0(j)k_0(n-j-1)$は，A がおこなうj字の返読は，< >内の括弧表示のパターン数$k_0(j)$と A< >以外の残りの$n-j-1$字の括弧表示のパターン数$k_0(n-j-1)$との積の数だけ繰り返すことを意味している．

次の$H_0(j)k_0(n-j-1)$は，j字括弧表示のなかで生じる返読は、 $n-j-1$字括弧表示のパターン数だけ繰り返すことを示している．

同様に，$H_0(n-j-1)k_0(j)$は，$n-j-1$字括弧表示のなかで生じる返読は，j字括弧表示のパターン数だけ繰り返すことを示している．

ハイフンの使用を 0 個または 1 個まで認める括弧表示の返読量の和$H_1(n)$も，括弧表示の先頭の 2 漢字を AB 型，A<B 型，A−B 型に分類する．

AB 型では B 以下の$n-1$字に依拠して，返読量は$H_1(n-1)$.

A<B 型は$H_0(n)$と同様に，$\sum_{j=1}^{n-1}\{jk_1(j)k_1(n-j-1)+H_1(j)k_1(n-j-1)+H_1(n-j-1)k_1(j)\}$.

A−B<型は，< >内にj字があるとすると，残り$n-i-2$字で，竪点の積の法則[11]に注意して， $\sum_{j=1}^{n-2}\{2jk_1(j)k_1(n-2-j)+H_1(j)k_1(n-2-j)+H_1(n-2-j)k_1(j)\}$. よって，

$$H_1(n)=H_1(n-1)+\sum_{j=1}^{n-1}\{jk_1(j)k_1(n-1-j)+H_1(j)k_1(n-1-j)+H_1(n-1-j)k_1(j)\}$$

$$+\sum_{j=1}^{n-2}\{2jk_1(j)k_1(n-2-j)+H_1(j)k_1(n-2-j)+H_1(n-2-j)k_1(j)\} \quad \cdots(5.2).$$

$H_2(n)$も同様に AB 型・A<B 型・A−B<型に加え，A−B−C<型の 4 タイプの和になる．

一般にハイフンがm個以下のn字括弧表示の返読量の総和$H_m(n)$は，

$$H_m(n)=H_m(n-1)$$

$$+\sum_{i=1}^{m+1}\sum_{j=1}^{n-i}ijk_m(j)k_m(n-i-j)+H_m(j)\,k_m(n-i-j)+H_m(n-i-j)k_m(j) \quad \cdots(5.3).$$

となる．

６．ｎ次順列とｎ次Ｋ順列の交点数の総和・平均

以上の公式の理論値を一覧表にしておく．

表 6.1　n 次順列と n 次Ｋ順列の交点数の総和・平均

n^{*1}	3	4	5	6	7
$\mu_n{}^{*2}$	9/6=1.5	72/24=3	600/120=5	5400/720=7.5	52940/540=10.5
$H_0/k_0{}^{*3}$	7/5=1.4	37/14=2.643	176/42=4.19	794/132=6.015	3473/429=8.096
H_1/k_1	9/6=1.5	55/19=2.895	299/64=4.672	1525/225=6.778	7489/816=9.178
H_2/k_2	9/6=1.5	58/20=2.9	325/69=4.71	1707/248=6.883	8603/919=9.361
H_3/k_3	9/6=1.5	58/20=2.9	329/70=4.7	1741/253=6.881	8836/942=9.38
H_4/k_4	9/6=1.5	58/20=2.9	329/70=4.7	1746/254=6.874	8878/947=9.375

*1　$n=2$のとき，$S_2/2!=1/2=0.5$．またすべてのmで$H_m(2)/k_m(2)=1/2=0.5$.

*2　$\mu_n=S_n \div n!=n(n-1)/4$.

*3　H_0/k_0は$H_0(n)/k_0(n)$の略．H_1/k_1なども同様．

この表からは，nが 4 以上であれば，一般の順列の場合の平均μ_nよりも各H_m/k_mすなわち各n字括弧表示の平均のほうが小さくなっていることがわかる．言い換えれば，$n \geq 4$のとき,交点数（返読量）は，n字括弧表示のほうが，デタラメに数字を並べ替えたn次順列より概して少ない．複雑度は交点数（返読量）をnで割ったものであるので，複雑度においても，n字括弧表示はn次順列よりも概して小さい．

なお，上記の表の$n=5$のH_2/k_2，$n=6$のH_2/k_2，$n=7$のH_3/k_3以降は，それまでの単調増加から減少に転じている．なぜこのようになるのか，理由はわからない．

７．今後の課題

白文の漢文の自動訓読システムの開発にあたっては，多くの訓読文の集積はもちろん，本稿で述べたような訓読構造の解析のほか，漢文を構成する個々の漢字および漢字どうしの，位置・役割・意味・読み方を明らかにした大規模な蓄積が必要であろう．

同時に，これまでの数理言語学や統計学を超える新たな思考が求められるであろう．

K 順列や交点数の理論は，おそらく組ひも群や有限の位相数学（代数幾何学）に関係しており，$H_m(n)$の一般項や項比の算出をはじめとして，発展する余地が大いにある．

数学では未解決の問題を証明するために，分野を超えた理論を用いることがある．フェルマー予想は,無関係と思われていた楕円曲線論を使って,ワイルズによって証明された．

これからの数理漢文学は，豊かな感性と広く深い数学の知識を使って，既成の手法や考え方の殻を破ることから生まれるのではないだろうか．

[1] bcy10369@kwansei.ac.jp（2023 年 3 月まで）.

[2] 計量国語学会 2019 島野達雄・古田島洋介・湯城吉信・田村誠「漢文の複雑度について」予稿の「3.数学的な補足」の「3-3.2 つの順列の距離」と「3-4.順列の交点数の定義」．以下の脚注に示した文献（予稿）はすべて個人ホームページ・和算序林 wasanjyorin.com に掲載している．

[3] 計量国語学会 2019「漢文の複雑度について」予稿の「1.返読する漢字数＝交点数」.

[4] 第 26 回数学史研究発表会 2019「漢文訓読の複雑度について－歴史情報学の試み－」の「3.数学的な補足(3)順列間距離の性質」.

[5] 0!=1．$S_0 = S_1 = 0$とする．$(1,2),(1,3)$は，$n \geq 0$ のとき成り立つ．μ_nはn次順列の交点数を昇順に並べたときの中央値であろう（未証明）.

[6] 応用数学合同研究集会 2012「漢文訓読の代数的構造」予稿の「1.漢文訓読とカッコ・竪点表示［定義 A]」.

[7] 計量国語学会 2012「漢文訓読における返り点・竪点システムの数学的構造について」予稿の「3.返読の基本構造」.

[8] [2]に同じ.

[9] 計量国語学会 2009「漢文訓読の数学モデル」予稿の「3.訓読漸化式と一般解」の「3.3 単返接モデルの一般項と極限値の意味」．OEIS（the Online Encyclopedia of Integer Sequence）の A000108.

[10] 計量国語学会 2016「数理漢文学への道(2)－線形と非線形のはざまで－」の付録として［基本データ集］に再読現象を含めたすべての「漢文数列」の母関数の関数 2 次方程式と線形微分方程式・線形差分方程式ほかのデータを掲載している．$k(n)$は，OEIS の A078482.

[11] 計量国語学会 2019「漢文の複雑度について」予稿の「2.重要な注意事項」の「2-2.交点数に関する積の法則」.

提出後追加

$\alpha = \alpha_1\alpha_2 \cdots \alpha_n, \beta = \beta_1\beta_2 \cdots \beta_n$のとき，

$\alpha(i) = j \Leftrightarrow \alpha^{-1}(j) = i$，$d(\alpha, \varepsilon) = d(\varepsilon, \alpha^{-1})$の証明.

αにおける要素iの位置（左から何番目か）がjのとき$\alpha(i) = j$と書く．このとき$\alpha_j = i$.
$\alpha * \beta = \beta_{\alpha_1}\beta_{\alpha_2} \cdots \beta_{\alpha_n}$と定義する．$\alpha * \beta = \varepsilon$のとき，$\beta = \alpha^{-1}$と書く．このとき$\alpha * \beta$の左から$j$番目の要素$\beta_{\alpha_j}$は$\varepsilon$の左から$j$番目の要素に一致するので，$\beta_{\alpha_j} = \beta_i = \varepsilon(j) = j$より$\alpha^{-1}(j) = i$.
$\begin{pmatrix}\alpha \\ \varepsilon\end{pmatrix}$のように上段に$\alpha$，下段に$\varepsilon$を置き，それぞれ右から$\alpha^{-1}$をかけると，$\begin{pmatrix}\alpha * \alpha^{-1} = \varepsilon \\ \varepsilon * \alpha^{-1} = \alpha^{-1}\end{pmatrix}$となる．上段では，$\alpha$の$j$番目の要素$i$は，$\varepsilon$の$j$番目の要素$j$に書き換わる．下段では，$\varepsilon$の$i$番目の要素$i$は，$\alpha^{-1}$の$i$番目の要素$j$に書き換わる．すなわち$\begin{pmatrix}\alpha \\ \varepsilon\end{pmatrix} \to \begin{pmatrix}\varepsilon \\ \alpha^{-1}\end{pmatrix}$の変換によって上段どうし下段どうしで同じ位置の要素iは，要素jに書き換わる．交点数は同じ数字どうしを線分でつないで求めるので，$d(\alpha, \varepsilon) = d(\varepsilon, \alpha^{-1})$が成り立つ.