◎漢文訓読の複雑度について ―歴史情報学の試み―

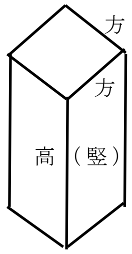
島野達雄（近畿和算ゼミナール）

１．関孝和と今村知商の漢文

関孝和の「求積」は，一組の向き合う二面が正方形である「直方体」の体積を次のように求めている.「有㆓方堡壔㆒(3)」等は, 有より先に方堡壔の 3 字を返読すること等を示す.

「仮如有㆓方堡壔㆒(3)，方面七寸高一尺二寸．問㆑積(1)」.「答曰積五百八十八寸」.「術曰置㆓方面七寸㆒(4)，自乗，得㆓四十九寸㆒(4)，又以㆓高一尺二寸㆒(5)相乗．得㆑積(1)」.

この漢文の漢字は 51 字．方堡壔のように返読する（先に読む）漢字の合計は 18 字で，返読率は 18/51=0.353 になっている.



同じ立体図形に対し，今村知商の『竪亥録』は，「今有㆓方竪之方竪㆒(5)，知㆑坪(1)式者」,「自因相因」.「以㆓方之尺数㆒(4)，自因乗，而得㆓歩数㆒(2)，于㆑是(1)，用㆓竪之尺数㆒(4)，因乗，則得㆓坪数㆒(2)．是寸坪也」であり，漢字 44 字．返読する漢字 19 字．返読率 19/44=0.432 になっている.

この他にも，二人が各々自著に書いた序や弟子の本に寄せた跋の返読率は，関よりも今村のほうが高く，関孝和の漢文より今村知商の漢文のほうが「複雑」と言えるだろう.

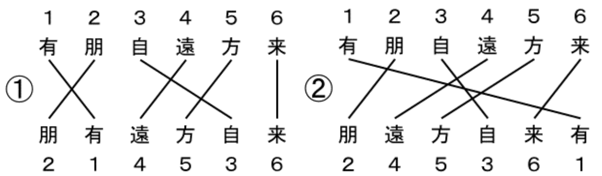
1.1 関孝和と今村知商が書いた漢文の返読率

関孝和	漢字数	返読数	返読率	今村知商	漢字数	返読数	返読率
求積・方堡壔	51	18	0.353	竪亥録・方竪	44	19	0.432
発微算法序文	181	103	0.569	竪亥録序文	389	285	0.730
発微算法演段諺解跋文	120	65	0.550	竪亥録仮名抄跋文	181	103	0.569

2．返読する漢字数＝交点数

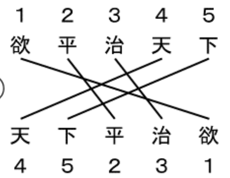
返読する漢字数は,「交点数」で可視化できる．①入れ子のない返り点，②入れ子をもつ返り点，③返り点・竪点，④返り点・再読，を認める場合の四つに分けて説明する.

有朋自遠方来という中国語に対し，①有㆑朋自㆓遠方㆒来（日本語の漢字列は朋有遠方自来)，②有㆘朋自㆓遠方㆒来㆖（日本語の漢字列は朋遠方自来有）の二通りの読み方がある．元の漢字列に 123456 と番号を付けると，日本語の漢字列は①214536，②245361 という順列（K 順列とよぶ）になる．上段に元の中国語と番号，下段に日本語の漢字列と K 順列を置き, 同じ漢字どうしを線分で結ぶと, ①は交点 3, ②は交点 7 を生じる (交点数とよぶ). ①②の交点数は，それぞれの返読する漢字数

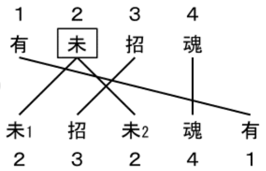


に一致している.

欲平治天下という中国語を③欲三平二‐治天下一（日本語の漢字列は天下平治欲）と一二三点と竪点を使って読んだ場合，K 順列は 45231，平と治がそれぞれ天下の 2 字を返読し，欲が平治天下の 4 字を返読するので，返読する漢字数は$2 \times 2 + 4 = 8$となり，この場合も交点数と一致する.

③ 

有未招魂という中国語を④有二$\boxed{\text{未}}$レ招魂一（日本語の漢字列は未招未魂有）と未を再読した場合，1234 は 23241 という数字列になる．23241 は 2 が 2 回あらわれているので順列とは言えない．ここでは，返り点・竪点に再読する場合をふくめた数字列を S 拡張順列とよぶ. S 拡張順列 23241 では，再読にともなう交点 1 が$\boxed{\text{未}}$で生じ，その他の交点 5 とあわせて交点数 6 となる.

④ 

交点数すなわち返読する漢字数を，元の中国語の漢字数で割った値を返読率とよぶ.

それぞれの返読率は，①3/6＝0.5，②7/6＝1.167，③8/5＝1.6，④6/4＝1.5 となり，①＜②＜④＜③の順に「複雑」になっていると言える.

3. 数学的な補足

(1) 順列と並べ替え

$1,2,\cdots,n$の n 個の数字を並べた数字列を n 次順列とよび，各数字を要素とよぶ.

$\alpha = e_1 e_2 \cdots e_n$のとき$e_n \cdots e_2 e_1$をαの逆順とよび，$\alpha^G$と書く．明らかに$(\alpha^G)^G = \alpha$.

$\alpha = e_1 e_2 \cdots e_n$に対し，$\rho = f_1 f_2 \cdots f_n$によって$e_i$が$f_i$番目の位置に並べ替わり、$\beta = g_1 g_2 \cdots g_n$となる$(g_{f_i} = e_i)$とき，$e_1 e_2 \cdots e_n \xrightarrow{f_1 f_2 \cdots f_n} g_1 g_2 \cdots g_n$, $\alpha \xrightarrow{\rho} \beta$, $\rho(\alpha) = \beta$と書く. $\varepsilon = 12 \cdots n$とすると$\alpha(\alpha) = \varepsilon$.

$\alpha \xrightarrow{\rho} \beta$かつ $\beta \xrightarrow{\sigma} \alpha$となるとき，$\rho$と$\sigma$は互いに逆順列であると言い，$\sigma^{-1} = \rho$, $\rho^{-1} = \sigma$と書く.

ρのi番目の要素をk_iとすると，σのk_i番目の要素がiとなっている．逆も同じ.

$(\sigma^{-1})^{-1} = \sigma$, $(\rho^{-1})^{-1} = \rho$, $\alpha^{-1}(\varepsilon) = \alpha$, $\alpha(\varepsilon) = \alpha^{-1}$, $\alpha^{-1}(\alpha^{-1}) = \varepsilon$, $\alpha(\alpha^{-1}) = \alpha$, $\alpha^{-1}(\alpha) = \alpha^{-1}$. また，$\varepsilon(\varepsilon) = \varepsilon$から$\varepsilon^{-1} = \varepsilon$, $\varepsilon^G(\varepsilon) = \varepsilon^G$から$(\varepsilon^G)^{-1} = \varepsilon^G$が成り立つ.

(2) 2 つの順列の距離

二つの n 次順列$\alpha = e_1 e_2 \cdots e_n$, $\beta = f_1 f_2 \cdots f_n$の順列間距離$d(\alpha, \beta)$を次のように定義する.

αのi番目の要素e_iとβのj番目の要素f_jが一致した（$e_i = f_j$）とする．このとき，e_iの右側にある要素の集合$\{e_{i+1}, e_{i+2}, \cdots, e_n\}$と$f_j$の左側にある要素の集合$\{f_1, f_2, \cdots, f_{j-1}\}$の共通集合の要素の個数を$d(\alpha, \beta, e_i)$とする．すなわち，$d(\alpha, \beta, e_i) = \left|\{e_{i+1}, e_{i+2}, \cdots, e_n\} \cap \{f_1, f_2, \cdots, f_{j-1}\}\right|$とする.

これは，αを上段に，βを下段に置き，要素$e_i = f_j$について，線分$e_i - f_j$を横切って交点を

生じる，上段のe_iの右側にある要素と下段のf_jの左側にある要素の，同じ要素どうしを結んだ線分の数をあらわしている．ただし，線分 2 本で 1 交点を生じる，と数える．つまり 3 本以上の線分は 1 点で交わらない，とする．こうして求めた$d(\alpha,\beta,e_i)$のすべての要素についての総和をα,βの順列間距離とよぶ．すなわち，$d(\alpha,\beta)=\sum_{i=1}^{n} d(\alpha,\beta,i)$と定義する．

この定義で，「上段のe_iの『左側』にある要素の集合」と「下段のf_jの『右側』にある要素の集合」としても，順列間距離は変わらないので，$d(\alpha,\beta)=d(\beta,\alpha)$が成り立つ．

(3) 順列間距離の性質

①任意のαについて，$d(\alpha,\varepsilon)=d(\varepsilon,\alpha^{-1})$.

$\alpha=e_1e_2\cdots e_n$とする．αとεにおいて，e_i（αのi番目の要素)をi に書きかえると，αはεに書きかわり，ε は，e_i番目の要素がiであるα^{-1}に書きかわる．このように書きかえても，同じ数字どうしを線分で結ぶので，$d(\alpha,\varepsilon)=d(\varepsilon,\alpha^{-1})$が成り立つ．

②任意のαについて，$d(\alpha,\alpha^G)=\frac{n(n-1)}{2}$.

二つの n 次順列ρ,σを上下段に置き，線分$i-i$と$j-j$が交点をもつとき，$c(\rho,\sigma)=1$と書き，交点をもたないとき，$c(\rho,\sigma)=0$と書く．このように決めると，αにおいてjがiの左側（L）にあればα^Gではjはiの右側（R）にあり，線分$i-i$と$j-j$は交点を生じるので，$c(\alpha,\alpha^G)=1$．この場合は$d(\alpha,\beta,j)$が 1 増える．.

α	L	R
α^G	R	L
$c(\alpha,\alpha^G)$	1	1

逆に，αにおいて，jがiの右側（R）にあれば，α^Gではjはiの左側（L）にあり，線分$i-i$と$j-j$は交点を生じるので，$c(\alpha,\alpha^G)=1$．この場合は$d(\alpha,\beta,i)$が 1 増える．
いずれにせよ 2 つの要素$i,j(i\neq j)$は交点 1 を生じる．交点の総数は，n 個の要素から 2 個の要素を組合せる数 ${}_nC_2$=n(n－1)/2. これは 2 つの n 次順列の順列間距離の最大値でもある．

③ 任意のα,βについて，$d(\alpha,\beta)+d(\beta,\alpha^G)=\frac{n(n-1)}{2}$.

α	L	L	R	R
β	L	R	L	R
α^G	R	R	L	L
$c(\alpha,\beta)$	0	1	1	0
$c(\beta,\alpha^G)$	1	0	0	1
$c(\alpha,\beta)+c(\beta,\alpha^G)$	1	1	1	1

同様にα,β の 2 つの要素$i,j(i\neq j)$の位置関係を考える．

αにおいてjはiの左側（L）か右側（R）のどちらかにある．βでもjはiの左側（L）か右側（R）のどちらかにある．α^Gではjはiの，αと反対側にある．

左の表から，すべての要素$i,j(i\neq j)$について，$c(\alpha,\beta)+c(\beta,\alpha^G)=1$となることがわかる．

$i,j(i \neq j)$の組合せは n(n－1)/2 通りあるので，$d(\alpha,\beta)+d(\beta,\alpha^G)={}_nC_2$=n(n－1)/2 をえる．

④任意のα,β,γについて，$d(\alpha,\beta)+d(\beta,\gamma) \geq d(\gamma,\alpha)$が成り立つ．

3 つの順列α,β,γにおいてiに対するjの位置は，全部で 8 通り．

この表から，すべてのi,jについて，$c(\alpha,\beta)+c(\beta,\gamma) \geq c(\gamma,\alpha)$となることがわかる．よって，三角不等式$d(\alpha,\beta)+d(\beta,\gamma) \geq d(\gamma,\alpha)$が成り立つ．

α	L	L	L	L	R	R	R	R
β	L	L	R	R	L	L	R	R
γ	L	R	L	R	L	R	L	R
$c(\alpha,\beta)$	0	0	1	1	1	1	0	0
$c(\beta,\gamma)$	0	1	1	0	0	1	1	0
$c(\alpha,\beta)+c(\beta,\gamma)$	0	1	2	1	1	2	1	0
$c(\gamma,\alpha)$	0	1	0	1	1	0	1	0

$d(\alpha,\beta)=d(\beta,\alpha) \geq 0$，$d(\alpha,\beta)=0$となるのは$\alpha=\beta$のときに限られるから，$d(\alpha,\beta)$は距離の 3 条件を満たしている．つまり，$d(\alpha,\beta)$を順列間距離とよんでもよいことがわかる．

4．複雑度を示す特徴量

⑴K 順列と交点数

２.①②③で K 順列αの交点数とよんだものは，とくに単位順列$\varepsilon=12\cdots n$とαの順列間距離$d(\alpha,\varepsilon)=d(\varepsilon,\alpha)$を指す．中国語の有朋自遠方来（123456）を朋有遠方自来（朋有り遠方より来る．K 順列 214536）に並べ替える順列は，$123456 \xrightarrow{215346} 214536$だから 215346 になる．K 順列 214536 は，有朋自遠方来を朋有遠方自来に並べ替える順列 215346 の逆順列（$215346^{-1}=214536$）である．すべての n 次順列αについて，$d(\alpha,\varepsilon)=d(\varepsilon,\alpha^{-1})$が成り立つ（３.(3)①）ので，有朋自遠方来を朋有遠方自来に並べ替える（およびその逆の）ときの交点の数を「朋有遠方自来（朋有り遠方より来る）の交点数」とよんでもよいことがわかる．

⑵歴史情報学への応用

このように漢文訓読の伝統を情報科学化（？）すると，豊富な研究教育資源である漢文史料を，歴史情報学（または歴史情報科学）の対象としてあつかうことができる．

初めに示したように，関孝和と今村知商の漢文の違いを，返読する漢字数つまり交点数や返読率によって数値化し，歴史上の人物ごとの訓読技法の特徴量としてあつかえる．

むろん和算にとどまらず，たとえば，李白と杜甫の詩の訓読，林羅山と渋沢栄一の論語の読み方の違いなど，著者別・年代別の比較や統計分析が容易におこなえる．

⑶K 順列の拡張と交点数

交点数の概念が，再読文字をふくむ場合にも拡張できることは２.④で述べた．n 次 S 拡張順列は，同じ要素が二度あらわれることもある長さ m の n 次 K 順列を拡張した数字列として定義でき，交点数と返読率も定義できる．

一般に，1,2,⋯,n の n 個の数字を，それぞれ重複を許して全部で m 個並べた数字列を，大きさmの n 次 Q 拡張順列とよぶ．

n 次 Q 拡張順列$\alpha_{n,m}$（大きさ m）と p 次 Q 拡張順列$\beta_{p,q}$（大きさ q）を考え，同じ数字（数字 1 から，数字 n と数字 p の小さいほうまで）どうしを線分でつなぐと，$\alpha_{n,m}$と$\beta_{p,q}$の「交点数」が求められる．$\alpha_{n,m}$を A 言語の単語・記号列（文），$\beta_{p,q}$を A 言語を B 言語に翻訳した単語・記号列（文）と読みかえると，得られた「交点数」[i]は，A 言語を B 言語に翻訳したときに生じる複雑さを示す特徴量のひとつと言える．

交点数・返読率の理論は，漢文訓読にとどまらず，英日，中英などの二か国語翻訳における語順変換の複雑度を測る「ものさし」となる可能性をもっている．

（以上）

---

[i] $\alpha = 123, \beta = 12, \gamma = 312$のとき，$d(\alpha,\beta) + d(\beta,\gamma) < d(\gamma,\alpha)$となり，三角不等式を満たさないので，Q 拡張順列間の「交点数」は距離とは呼べない．