

1. 接続、返接、連結記号の定義

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、漢文言語 Σ^* の元を「漢字列」とよぶ。

$$\Sigma^* = \{\lambda, \sigma_1, \sigma_2, \dots, \sigma_n, \sigma_1\sigma_1, \sigma_1\sigma_2, \dots, \sigma_1\sigma_n, \sigma_2\sigma_1, \dots, \sigma_8\sigma_1\sigma_1\sigma_3\sigma_1\sigma_5, \dots\}$$

ここで、 λ は空順列を指し、「空の漢字列」とよぶ。

Σ^* の元 u, v に対して、 u, v を並べる二項演算を「接続」とよぶ。 $u \cdot v = uv$

また Σ^* の元 u, v に対して、 u, v の順序を入れ替える二項演算を「返接」とよぶ。 $u \langle v \rangle = vu$

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 \langle を用いて $u \langle v \rangle$ 次のように表記する。 $u \langle v \rangle = \sigma_3 - \sigma_1 - \sigma_3 - \sigma_2 \langle v \rangle$

以下、接続演算の \cdot 記号を省略し、接続と返接が続くときは返接演算が優先するとする。

2. カッコつき表示の定義

漢文言語 Σ^* において、「カッコつき表示」を次のように再帰的に定義する。

(i) 漢字 1 字はカッコつき表示である。

(ii) 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。

(iii) P を漢字 1 字または連結記号で結ばれた 2 字以上の漢字、 m をカッコつき表示とするとき、 P と m の返接 $P \langle m \rangle$ はカッコつき表示である。〔注〕参照。

3. n 個の異なる漢字の接続とカッコつき表示

n 個の異なる漢字の接続を K_n と書く。

K_n に対して、漢字の順番を変えることなく、連結記号 \cdot 、カッコ開始記号 \langle 、カッコ終了記号 \rangle の記号を挿入したカッコつき表示全体の集合を $[K_n]$ と書き、「 K_n から生成されるカッコつき表示」とよぶ。

$[K_n]$ 、 $[K_m]$ に対して、集合の接続、返接を次のように定義する。

$$[K_n][K_m] = \{uv \mid u \in [K_n], v \in [K_m]\}$$

$$[K_n] \langle [K_m] \rangle = \{u \langle v \rangle \mid u \in [K_n], v \in [K_m]\}$$

$[K_n]$ の元の総数を a_n であらわす。 $a_1 = 1$ 、 $a_2 = 2$ は明らか。

なお、 K_0 は空の漢字列 λ とし、便宜上 $[K_0] = \{\lambda\}$ 、 $a_0 = 1$ と定義する。

4. K_n から生成されるカッコつき表示の総数 a_n の漸化式

以上の定義から、

$$a_n = 2a_{n-1} + \sum_{i=1}^{n-2} a_i a_{n-i-1} + \sum_{\substack{i+j=1 \\ i \geq 1, j \geq 0}}^{n-2} a_i a_j \quad \text{すなわち} \quad a_{n+1} = 3a_n + a_{n-1} + \sum_{i=2}^{n-2} a_i a_{n-i} \quad (n \geq 4) \quad \text{を得る。}$$

この漸化式を逐次計算すると、たとえば、「14 字の異なる漢字からなる漢文(中国語)が与えられたとき、日本語の語順として成り立つ可能性があるのは、並び替えの全体(14!)の 0.00017 程度しかない」ことがわかる。

〔注〕

カッコつき表示の定義 (iii) で「 P を漢字1字または連結記号で結ばれた2字以上の漢字」とし、訓読の木の定義(8)で「返節の左辺の下位に返節があらわれない」としたのは、ふつう漢文(中国語)では、「返って読む部分」に対する「返って読まれる部分」は、漢字1字または連結記号で結ばれた2字以上の漢字になっていることに対応している。

つまり、漢文(中国語)から日本語への翻訳(ここでは漢字の語順変換)をおこなうメタ言語としての返り点・連結記号システムを形式化するにあたり、返接 $P\langle m \rangle$ において、 m は何重でも入れ子構造と並置構造をとりえるように約束し、 P は「漢字1字または連結記号で結ばれた2字以上の漢字」と約束したほうが実情に即している。

ただし、この P をより広く「カッコつき表示」と定義することもできる。

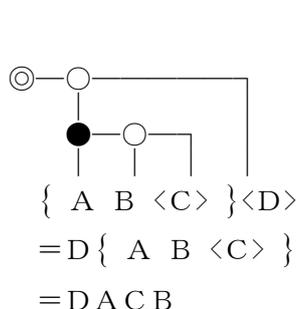
このばあい、演算の優先順位を示す $\{ \}$ 記号などを用いる必要がある。

たとえば、異なる4字の漢字からなる K_4 では、 $ABCD$ という漢字列から、 $BDAC$ 、 $CADB$ 、 $DACB$ 、 $DBAC$ の4つを生成することはできない($a_4 = 4! - 4 = 20$)。

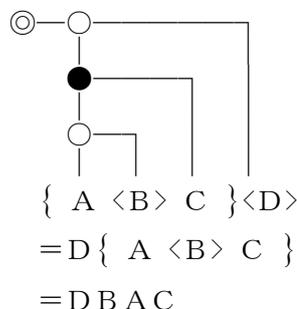
しかしこのうち $DACB$ と $DBAC$ は、 $\{ \}$ 記号を使えば、 $\{ AB\langle C \rangle \}\langle D \rangle = DACB$ 、 $\{ A\langle B \rangle C \}\langle D \rangle = DBAC$ と生成することができる。

同じように、訓読の木の定義(8)で「返節の下位に返節があらわれてもよい」、言いかえれば「返節の左辺は、下端に行き止まりの漢字、連節のどちらかを持ち、返節をもたない」だけにすると、 $DACB$ と $DBAC$ は下図のようにグラフ化できる。

$ABCD \Rightarrow DACB$ の訓読の木



$ABCD \Rightarrow DBAC$ の訓読の木



漢文訓読の実情に即した狭義の定義に対して、このように $\{ \}$ 記号の使用を認めた定義を「広義のカッコつき表示」、「広義の訓読の木」とよぶ。

まだ正確な判断とは言えないが、漢文(中国語)から英語への翻訳(ここでは語順の変換)では、広義のカッコつき表示を採用したほうがよいように思う。

『論語』の「有朋自遠方来、不亦乐乎」は次のように英訳されている(Legge 1861)。

Is it (乎) not (不) delightful (楽) to have (有) friends (朋) coming (来) from (自) distant (遠) quarters (方)? (乎)

この英訳の構造は、広義のカッコつき表示をつかえば表現できる。

$\{ \text{有朋自} - \text{遠} - \text{方} \langle \text{来} \rangle \} \langle \text{不} - \text{亦} - \text{楽} \langle \text{乎} \rangle \rangle$

ここで、Is it...?は疑問反語の漢字「乎」を再読している、とも解釈できる。toは訓読の送り仮名「コト」とみなせ、漢字「亦」は黙字(読まない助辞)と言えないこともない。