

7J-6

日本語プログラミングにおける  
識別子の分析とモデル作成

尾関 哲<sup>1</sup> 佐藤 邦弘<sup>2</sup> 太田 健一<sup>3</sup> 宮脇 富士夫<sup>4</sup>  
 神戸高専<sup>1</sup> 姫路工業大学<sup>2</sup>

1. はじめに  
 我々は、プログラムの生産性を向上させる手段として日本語プログラミングに着目し、図1に示す開発環境を構築した。本環境のエディタは、日本語識別子の簡略入力方式に特徴がある。その評価を行なうためには日本語プログラミングにおける識別子のモデルが必要となる。しかしながら、現状では実用的な日本語プログラムが十分存在しない。そこで、我々が本環境の中で開発した日本語プログラム<sup>1)</sup>の中から識別子を抽出し、識別子のモデルを作成したので報告する。

日本語化し、識別子に日本語が使えるようにしたものである。ここで識別子とは、変数名、関数名、定数名、マクロ名などを指す。入力効率には入力モードの切り替え、カナ漢字変換が影響することから識別子を漢字のみ、*カカナ*のみ、ASCII文字のみ、その他各種の文字の混合した文字列の4グループに分類した。その分類結果を表1に示す。また、入力時のキータッチ回数は、同一グループ内での識別子文字列の長さ(バイト数)に比例すると考えられるので、各グループについてその度数分布を取った(図2~5)。

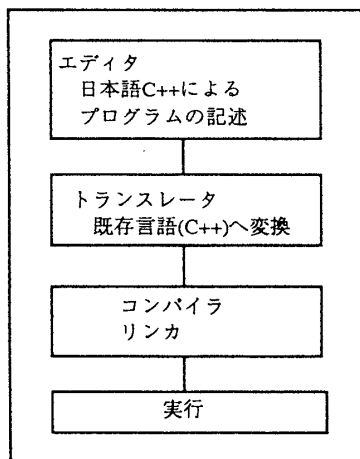


図1 日本語プログラミング環境

2. 識別子の抽出

日本語プログラミング用エディタの評価を行なうためには実際のプログラム中で使用される識別子のモデルが必要である。そこで、まず日本語C++で作成したプログラム(約3600行)から識別子を抽出した。日本語C++とは、C++の予約語を

表1 種別の割合

	種類	割合(%)
漢字	212	47
カカナ	17	4
ASCII	38	8
混合	188	41
合計	455	100

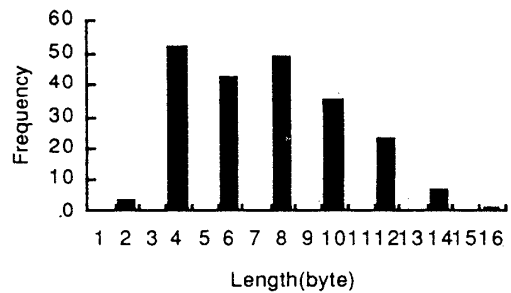


図2 漢字グループ

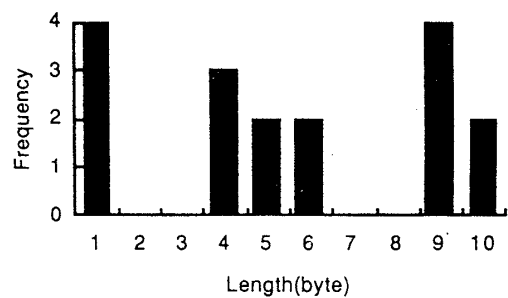


図3 カカナグループ

Analysis of Identifiers in Japanese-based Programming and its Model

Satoshi OSEKI<sup>1</sup>, Kunihiro SATO<sup>2</sup>, Kenichi OHTA<sup>3</sup>,  
 Fujio MIYAWAKI<sup>4</sup>

Kobe City College of Technology<sup>1</sup>  
 Himeji Institute of Technology<sup>2</sup>

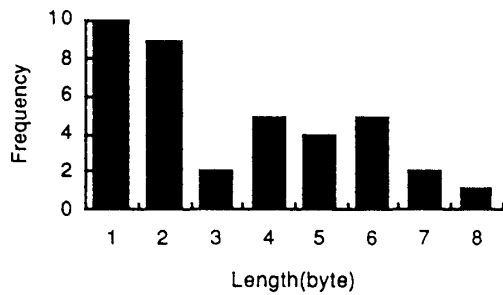


図4 ASCIIグループ

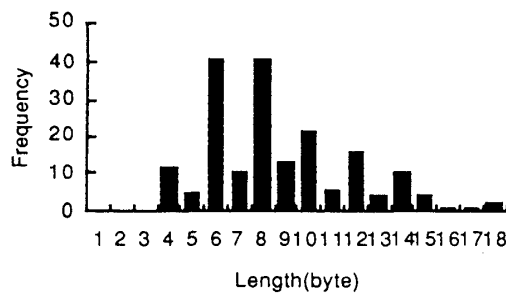


図5 混合グループ

### 3. 識別子モデル

抽出した識別子の統計結果をもとにエディタの評価に使用するための識別子のモデルを作成した。以下にモデルの作成手順を述べる。まず、取り出す個数を表1の種別割合で分割する。取り出す識別子の数をNとすると、表1より、漢字グループから  $N \times 47/100$ 、カタカナグループから  $N \times 4/100$ 、ASCIIグループから  $N \times 8/100$ 、混合グループから  $N \times 41/100$  のように分配する。つぎに各種別毎に文字列の長さ分布(図2~5)に応じて取り出す。今回、この手順で50個、100個、200個、300個、400個の識別子モデルを作成した。識別子数100個の場合のモデルを表2に示す。

表2 識別子のモデル

陽	入力最終行	「」の数
目印	演算子情報	キ入力
黄色	名前要素数	コール_2
仮想	色画面表示	左上_X
本体	直前比較名	リスト表示
移動	演算子印刷	1行印刷
関数	被置換名要素	関数名_2
変位	関数頭部処理	文字列_1
属性	全名前文字数	定義モード

限界	要素位置調整	ファイルの数
先頭	関数頭部情報	要素長_2
自動	関数階層順先頭	桁位置_1
下移動	全演算子文字数	ファイル追加
要素長	」	指示子_1
終文字	コール	命令コード
着目行	レジスタ	1文字読込
桁位置	ファイル名*ラ	関数名リスト
表示色	{	1行削除_1
入力順	h	名前情報_1
行接続	::	行番号まで
行検索	cx	コメント文字数
名前置換	far	画面1行クリア
演算子表	NULL	表示文字_1
表示位置	SREGS	終わり桁マーク
現在時刻	time_t	演算子番号_2
位置検出	桁_4	次非コメント要素
符号記号	行_6	関数宣言子_1
選択番号	桁_2	宣言指定子_1
単引用符	リスト済	1文字読み込み
位置検索	文字_3	移動または複写
定数検索	カウンタ_5	ウィンドウ部分消去
画面再現	無用_1	日本語プロパティ
名前開始行	入力_1	
定数文字数	右下_X	

### 4. おわりに

日本語プログラミング言語の普及のためには、言語自体の特質、開発環境に対する考察が重要である。今回は、開発環境の一部であるエディタの評価の材料としての識別子モデルを作成することができた。今後は、この識別子モデルを使った日本語エディタの評価方法を検討し、我々の提案する簡略入力方式の有用性を検証する予定である。

### 参考文献

[1] 尾関, 佐藤, 宮脇: 日本語識別名入力機能を強化した日本語プログラミング用エディタの開発, 第8回ソフトウェアコンファレンス(1992)