

オブジェクト指向と名前管理に基づく

ユーザインタフェース

-SIMPOS・Smalltalkにおける名前づけ-

青柳龍也・田中英彦

(東大工学部)

6U-4

1. はじめに

我々は、個人向きオブジェクト指向計算機システムORAG A[1]の、名前管理を行うモジュールNameMaster[2]の開発を行っている。NameMasterは、名前に関するデータベースを保持し、名前に関する種々のサービスを提供する。最も大事なサービスは、新しくシステムに導入される名前の記号性の保持である。NameMasterは、すでに記号性のある名前づけがなされている、かなり大きなライブラリの存在を仮定している。そのようなライブラリ中の名前をもとに新しい名前の記号性をチェックする。

このため、NameMasterの開発には、すでにあるかなり大きなライブラリ中の名前づけを調べる必要がある。今回は、ICOTの開発した推論マシンPSI上のOS、SIMPOS[3]と、Xerox社のSmalltalkシステム[4]について、クラス名のつけ方を調査した。なお、SIMPOSはバージョン2.52、Smalltalkはバージョン2である。

以下、次節で記号性をチェックするNameMasterの構成について簡単に述べ、その後、調査の結果を報告する。

2. 記号性をチェックするNameMaster

記号性をチェックするNameMasterには、1)基本単語解析、2)構文解析、3)記号性のチェックの三つのフェーズがある。これは、一般のコンパイラが1)字句解析、2)構文解析、3)中間コード生成(意味づけ)の三つのフェーズを持つことに対比させることができる。

基本単語解析は、基本単語辞書を用いて名前を構成する基本単語を解析する。ここで、同意語、省略形の処理も行なわれる。

構文解析は、名前の構成規則に基づき名前の構文を解析する。名前の構成規則は、名前に許される文法である。コ

表1 語彙の大きさ

	SIMPOS	Smalltalk
単語	3948	444
見出し語	697	183
英単語	575	179
非英単語	122	4
共通見出し語	96	

ンパイラの場合には、生成される中間コード(意味)が構文によって決まるように、NameMasterでは使われた構成規則により名前の持つべき記号性が決まる。

基本単語解析では、どのような基本単語辞書を用意するかが問題となる。構文解析では、どのような構成規則を用いるかが問題となる。今回の調査は、この二点の参考とするために行った。

3. 基本単語辞書の調査

3.1 語彙

語彙の調査のためには、語彙の要素となる単語の認定法を決める必要がある。

SIMPOSおよびSmalltalkでは、単語は省略しないで使うことが一般的であり、単語と単語の境界も明示されている。SIMPOSでは「_」で区切られた文字列を、Smalltalkでは大文字で区切られたものを単語とした。

また、単語と見出し語の区別も必要である。異なるクラス名中で使われている同じ文字列は、単語としては別のものだが、見出し語としては、同一のものである。

表1に語彙の大きさを示す。英単語かどうかの判断にはUNIXのspellを用いた。SIMPOS、Smalltalkそれぞれに関して、使用頻度の多い見出し語を表2に示す。表3には、SIMPOSとSmalltalkに共通して使われている見出し語を示す。

表2 使用頻度の多い見出し語

SIMPOS		Smalltalk	
見出し語	頻度	見出し語	頻度
as	177	controller	26
with	146	view	22
window	143	change	16
file	139	list	12
tap	77	node	11
menu	69	text	11
editor	65	class	9
buffer	49	file	9
error	47	form	9
manipulator	47	method	9

見出し語の数はそれほど多くない。SIMPOSの全ソースコード(約75万単語)でも、見出し語は3500語程度である。また、ほとんどは英単語である。共通の見出し語の占める割合もSmalltalkでは50%を越えている。

今回集めた見出し語をもとに基本単語辞書を構築することが可能である。

表3 共通の見出し語

見出し語	SIMPOS		Smalltalk		平均%
	頻度	%	頻度	%	
file	139	3.5	9	2.0	2.8
change	1	0.03	16	3.6	1.8
list	30	0.8	12	2.7	1.7
menu	69	1.8	5	1.1	1.4
class	27	0.7	9	2.0	1.4
text	9	0.2	11	2.5	1.4
node	4	0.1	11	2.5	1.3
editor	65	1.7	3	0.7	1.2
method	10	0.3	9	2.0	1.1
display	17	0.4	7	1.6	1.0
:	:	:	:	:	:
合計	1094	27.9	247	55.6	41.7

3.2 クラス名の長さ

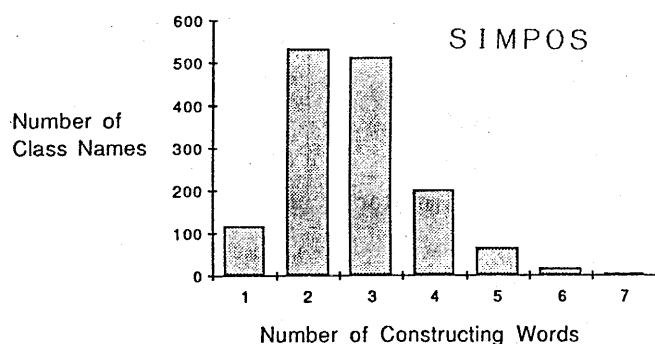
クラス名の長さに関する統計量を表4に示す。単語は省略しないという原則のため、名前の長さはかなり長い。このため、NameMasterは省略形をサポートする必要がある。SIMPOSの方が長いのは、`_`で単語を区切ることにも原因がある。

表4 クラス名の長さ

	SIMPOS	Smalltalk
平均	16.6	12.2
標準偏差	6.6	5.6
最大	47	31
最少	2	3 (文字数)

4. 名前の構成規則用の調査

4.1 名前を構成する単語数



一つのクラス名を構成する単語数の分布を図1に示す。

4語以上の名前は少ないことがわかる。名前の構成規則は3語程度に制限できそうである。

4.2 英語の品詞に基づく名前の構成

Smalltalkの全クラス名に関して、どのような品詞から構成されているかを調べた。結果を表4に示す。

Smalltalkのクラス名のほとんどは、少数の構成規則で作られていることがわかる。名詞のみでできている名前が75%を占める。4つ以上の名詞から構成されているクラス名は存在しない。また、最後の単語が名詞以外であるようなクラス名も存在しない。

表4 Smalltalkのクラス名の構成

	頻度	%
名詞+名詞	84	37.8
名詞	62	27.9
名詞+名詞+名詞	21	9.5
形容詞+名詞	9	4.1
動詞原形+名詞	8	3.6
過去分詞+名詞	6	2.7
形容詞+名詞+名詞	4	1.8
その他	28	12.6

5. おわりに

NameMasterの基本単語辞書と名前の構成規則を作るために、実際のオブジェクト指向システムの名前を調べた。今回の結果をもとに、NameMasterを設計していく予定である。

<参考文献>

- [1] 神田, "並列オブジェクト指向計算機システムORAGA", 東京大学情報工学博士論文, 1986
- [2] 青柳他, "NameMasterの実現に向けて", 情報処理学会第33回全国大会5D-4, 1986
- [3] 新世代コンピュータ技術開発機構, "SIMPOS総合説明書(第一版)", 1986
- [4] A. Goldberg and D. Robson, "Smalltalk-80 The Language and its Implementation", 1983

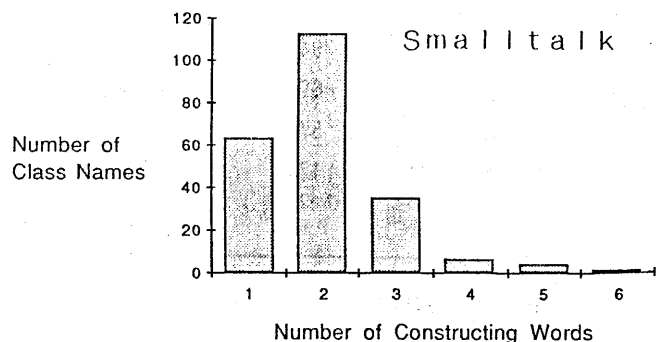


図1 クラス名を構成する単語数