

# 移動体関連 RFC の相互関係と翻訳のための辞書作成

6T-08

日榮 雄二(\*) 堀 考洋(\*) 小川 清(\*\*)  
大同工業大学(\*) 名古屋市工業研究所(\*\*)

## 1 はじめに

RFC(request for comment)は、インターネットの規約に関する情報源である。WEB で英語の全文が公開されている。同一の用語が、文脈上異なる意味を表す可能性があるため、RFC 中の用語を正確に理解するために、参照している他の RFC との関係、順に辿り、個々の単語の分析、熟語の分析および文脈の分析を行うことを検討した。まず、相互関係を自動的に収集するためのツールを作成し、関連づけを調べ、RFC 翻訳用辞書を作成してきた。

現在研究中の携帯(Mobile)関連の RFC について、単語抽出、個々の出現頻度を調査し、昨年度行なった RFC 全体との比較と、単語や用語の分析・検討を行った。

## 2 RFC 参照情報の取得

RFC では、原則として参照した RFC を最後の“References”で一覧している。一つの RFC ファイルを読み、その参照情報の RFC を識別し、参照情報としてデータを自動的に取得するソフトウェアの簡易版を作成した。

一つの RFC ファイルで、“References”まで読み飛ばし、その後ろにある RFC で始まる 4 桁の数字のみを読み取ることにした。参照情報として取得した RFC のファイルを読み、その参照情報を読み取る上位のソフトウェアを作成した。

ここでは、管理ファイルを作成し、そこで参照している RFC の 4 桁までの数字を、小さい方から順にならべることにした。RFC は 2000 年 12 月末現在 3030 件のみであるため、4 桁処理でも当面可とした。右辺に現れる回数が多いものは、基礎的なものか、多くの規約を利用した応用例である。

## 3 カタカナ語

ここでは、英語の発音をカタカナ表現したものを「カタカナ語」と呼ぶことにする。

通信規約は、専門家のみが読むことが想定されている場合には、英語のままに翻訳されない文書がある。翻訳された場合でも、専門用語は、カタカナ語または英単語のままのものもある。

Daido Institute of Technology(\*)

例えば、1 行の中に複数のカタカナ語があったり、3 語以上のカタカナ語から構成される複合語が出現する場合がある。

ハードウェア寄りの通信規約、数学的な通信規約、及び直接英単語で構成されるプログラミング言語で記述すればよいものは、英語のままの方が便利な場合もある。

住所、氏名などの固有名詞、文字コード、文字フォントのように自然言語に依存したものの場合、専門家以外に利害が及ぶ文書は翻訳されると、現在の課題に対する理解が広がることもある。

専門家の場合も、日本語で考える場合と、英語で考える場合とで異なる意味で捕らえている可能性もある。意味を厳密に理解したり、日本語で説明する場合に、日本語に翻訳すると、意味の重複、用語体系の構造について問題点が整理でき、英語だけで理解していた段階では考察しなかった知見を得ることもある。

通信規約上に実装する自然言語依存の部分は少ない可能性があるが、送受信する文字の制約に対する理解が、明確になることもある。

## 4 カタカナ語辞書

通信規約では、コンピュータ用語と同様に、接頭語、接尾語を加えた造語、名詞の動詞化の事例がある。これらは、文脈に沿って翻訳を見比べて比較することが重要である。

カタカナ語でこれまでの RFC、通信関連の専門用語辞書が作られている単語は、どのような漢字の単語または熟語が選択可能かを判断するには、日本語の用語の理解が必要である。

そのため、各種専門用語、各種カタカナ語辞書から、2 千語の通信関連のカタカナ語辞書を作成し、選択可能な用語の候補一覧として用いることとした。

また、翻訳家、言語学の専門家の協力を得るため、kata-kana@egroups.co.jp によりメーリングリストを運営し、各方面の専門家と議論を進めている。

Relation of RFC on Mobile and English-Japanese dictionary

HIEI Yuji(\*), HORI Takahiro(\*), OGAWA Kiyoshi

## 5 用語辞書

昨年作成した、TR、AWK及びC言語によるRFCから文字列を直接単語を抽出するツールでは、図形を表現するための記号として利用しているアルファベットを単語として抽出したため、新規語との区別を厳密につけることが困難であった。

本年は、特定の分野（移動体）のRFCを翻訳しながら、専門用語辞書の拡充と絞込みを行うことを目指した。昨年行なった、全文書の単語を抽出して辞書を作るという上から下への方法と、本行なっている1つ一つのRFCを翻訳し、その用語辞書をまとめるという下から上への方法の、両面から辞書作成を推進したことになる。

本年は、具体的には移動体に関するRFCを体系的に翻訳することにより、個々の用語の意味、関連する意味を検討した。具体的には、表1の3文書を基点として3段階まで関連性を調査した。

## 6 用語の検討

具体的な用語の検討の一例は、下記のとおりである。通信用語では、カタカナ語が広く使われている。それらが実際の文脈中でどのような意味合いを取っているかを確かめながら検討している。

**agent** : 代理人。携帯端末の代理として、ネットワーク側で携帯端末の役割を担う振舞いを行なう。何もしてくれないエージェントではなく、ネットワーク側からは携帯端末として見えるように働いている。擬人化して使われていると考え、代理ではなく代理人とした。

ただし、漢字の用語のみでは、英語との対応関係が不明確になるため、代理人(**agent**)という表記を取る。

## 7 用語の比較

RFC中の単語の抽出はTRコマンド、AWKスクリプト、Cプログラムを用いて作成し、単語の出現回数を集計する過程は、昨年の研究成果を用いた。

(1) 単語を小文字に統一し、数字、記号を削除した(ピリオドなど)。

(2) 同一の文字が3回以上連続する単語を削除した(ZZZなど)。

(3) 副詞(ly)、動名詞(ing)、複数形(s, es)、過去形(d, ed)は同一単語として集計した。作成した辞書は、表計算ソフトウェアで分析することとした。

## 8 用語の出現回数の検討

各RFCで、一般用語を削除し、専門用語のみの

優先順位付けをした用語一覧を作成した。RFC2002, RFC2003, RFC2004についての例は表1の通りである。

表1 RFC2002, RFC2003, RFC2004 専門用語頻度辞書

2002 用語	頻度	2003 用語	頻度	2004 用語	頻度
agent	732	ip	133	ip	67
mobile	704	datagram	97	header	48
node	597	encapsulator	62	encapsulation	28
Home	550	tunnel	59	datagram	26
registration	433	header	55	address	18
address	428	icmp	35	forwarding	15
foreign	292	source	32	source	12
ip	275	destination	31	protocol	10
request	211	sender	27	destination	9
mobility	198	encapsulation	25	rfc	9

参照関係の強いものは、相互の用語に類似度が高いことが想定された。RFC2003, RFC2004の間では、上位10語中に6語同じ用語があることが認められた。

昨年作成した全体の辞書に比べると、雑音(ノイズ)相当の図形を表すための文字列は少なく、それらを取り除く処理は必要なかった。

## 9 今後の課題

本研究では、RFCにおける用語辞書の作成とともに用語定義を行う上での資料の作成を進めた。英和辞書に搭載されていない単語、専門辞書に搭載されていない単語で、特定の専門用語としての意味の範囲、内容が、RFC全体の中でどのような限定がされているか、理解できていない単語もある。

そのため、各RFCの意味について理解し、同一分野の用語を個別に並べ、用語比較することにより、よりよい訳語を選択して分野別辞書の作成を行なう方向である。

## 参考文献

- [1] RFCを対象とした標準用語辞書の作成と分析  
平成11年電気関係学会東海支部, 小川清ほか
- [2] 不要語リストを用いたRFC英和辞書作成過程の課題, 言語処理学会第6回年次大会, 森理ほか