マニュアル概要例文間のファジー関係に基づく英日機械翻訳エンジン jfreshman (Japanese version of Fuzzy REtrival SHell for a command MAN) [1]

4K－2

安達　久博，下山　豪彦 [2]

宇都宮大学　工学部　情報工学科 [3]

1　はじめに

近年，用例に基づく機械翻訳システムの研究が各所で盛んに行われている．この翻訳方式は，長尾 [1] によって提案された「アナロジーによる翻訳」という考え方を基本にしている．この方式の翻訳戦略は，「ある文を翻訳するために，その文によく似た文の対訳用例を模倣利用して，翻訳文を得る」というものである．佐藤 [2] は，この方式の実現例として複数の翻訳用例を利用することで一文全体の翻訳を行う方式を提案している．これらのプロトタイプ・システムが利用する用例は，基本的に形態素解析処理されたコーパスを対象としている．従って，入力文も形態素処理する必要がある．また，類似用例を検索する際に品詞列や構文解析結果である文構造（例えば，木構造）の比較を利用している．

本稿が提案する方式の特徴は，辞書や翻訳規則を一切使用せずに，文字列形式の対訳用例のみを利用して，一文全体を翻訳するものである．入力原文は英語文のため空白等により単語単位に分割されてはいるが，翻訳処理の基本は，文字あるいは文字列処理を基本としている．

2　マニュアル概要文の特徴

本研究で利用する対訳用例，ならびに翻訳対象とするマニュアル概要文とは，図1に示したUNIX計算機上のオンラインマニュアルのコマンド名の機能を一行文で解説しているwhatisデータベースである．

```
bdiff(1) - diff for large files
bdiff(1) - ラージ・ファイルの diff
bifchmod(1) - change mode of a BIF file
bifchmod(1) - BIF ファイルのモード変更
biffind(1) - find files in a BIF system
biffind(1) - BIF システムにおけるファイルの検索
bifls(1) - list contents of BIF directories
bifls(1) - BIF ディレクトリ内容のリスト表示
bifmkdir(1) - make a BIF directory
bifmkdir(1) - BIF ディレクトリの作成
bifrm, bifrmdir(1) - remove BIF files or directories
bifrm, bifrmdir(1) - BIF ファイルまたはディレクトリの削除
cflow(1) - generate C flow graph
cflow(1) - C フロー・グラフの生成
```

図.1　whatis データベースの例

この概要文は「コマンドの機能を簡単に説明する」という目的から，文中で使用される語彙や構文パターンは一般の自然言語文に比べて非常に限定された範囲の文集合と捉えることができる．例えば，英語文の主語（コマンド名）は省略され，文の先頭が動詞で始まる文形式を採用している．また，対訳用例中では固有名詞の一部は日本語文中でも英語表記のまま出現している．これは，UNIXシステム固有の専門用語をあえて日本語訳に変換しなくても運用上問題が無いことを意味している．逆に，無理に日本語訳を生成することはユーザの解釈の妨げになると考えられる．

3　翻訳方式の概要

前章で述べたマニュアル概要文の特徴を考慮すると，翻訳処理は基本的に，核となる文構造の部分のみを翻訳対象とすれば良いことを意味する．翻訳方式は以下に示す2つの手順で行われる．

1. 最適な類似対訳例文の検索　　2. 原文と類似例文との差分（ギャップ）に対応する部分の検索

[1] English-Japanese Machine Translation Based on Fuzzy Relations between Example Sentences
[2] Hisahiro ADACHI and Hidehiko SHIMOYAMA
[3] Department of Information Science, Utsunomiya University

例えば，A社のコマンド **rmdel** の概要文「remove a delta from an SCCS file」を入力原文とすると，B社の対訳用例中で最も似ている概要文としてコマンド **colrm** の「remove columns from a file:ファイルからカラムを削除する」が検索される．英語側の共通部分は「remove X from Y file」となる．この時点では英語側と日本語側がどういう対応関係かは未定である．但し，原文との対応関係は「X=delta」,「Y=SCCS」は確定している．次に英語側と日本語側との対応関係を調べるため，対訳用例の文集合から「remove」のみを共通部分として含む文ペアを検索し，同時に日本語側の共通部分を抽出する．すると,「remove = を削除する」が得られる．同様にして,「from=から」,「file=ファイル」が得られ，英語側に対する日本語訳側の「Y ファイルから X を削除する」という対応関係が確定する．以上により最終的に「remove delta from SCCS file」=「SCCS ファイルから delta を削除する」という訳文が得られる．尚，文間の比較の際に冠詞は無視される．

4 文間の類似性を判断するファジーメンバーシップ関数

「A と B は等しい」というような明確な関係に対して「A と B はよく似ている」という様なあいまいな関係はファジー関係として定義できる [3]．つまり，対訳用例を全体集合 X とすると，直積 $X \times X$ におけるファジー関係 Rはメンバーシップ関数$\mu_R : X \times X \rightarrow [0,1]$ により特徴づけられたファジー集合 Rとなる．しかも，ファジー関係が反射律，対称律を満たす時，相似関係と呼ばれる．文間の相似関係を単語列の一致度を重視したメンバーシップ関数として以下のように定義した．

$$R(A,B) = \frac{LCS(A,B)}{Length(A) + Length(B) - LCS(A,B)} \quad (1)$$

ここで，関数 $Length(A)$ は文（単語列）A の単語数であり，関数 $LCS(A,B)$[4] は二つの文 A，B 間の最長共通部分単語列の数を計算する．従って，対訳用例から英語側と日本語側の単語の対応関係を調べる場合，$LCS(A,B) = 1$ となる文ペアを求めることを意味する．

5 実験と検討

NEWS-OS4.2C の whatis データ 160 文を対訳用例とし，HP-UX 9000 の英語 whatis データから 20 文を抽出し入力原文とした．その結果，翻訳成功率 90% を得た．ここで翻訳に失敗した文は対訳用例中で同一の英単語に対して複数の訳語（例えば，カタカナ表記と漢字表記）が対応し，英語側と日本語側の対応関係が同定できなかった場合である．今回の実験では，whatis データベースをそのまま利用したが翻訳成功率を向上するため，事前に同一単語に対する訳語の統一等の正規化処理を対訳用例に対して行う必要がある．

6 おわりに

本稿では，マニュアル概要文という文法的にも意味的にも非常に限定された領域の自然言語文を辞書や翻訳規則を一切利用せず，マニュアル概要文の対訳用例データベースだけを翻訳知識源として利用し，文間の相似関係をファジー関係のメンバーシップ関数で定義し，複数の対訳用例を組み合わせることで訳文を生成する英日翻訳処理方式を提案した [4]．

参 考 文 献

[1]Nagao,M:A Framework of Mechanical Translation between Japanese and English by Analogy Principle, in Artificial and Human Intelligence,Elithorn and Banerji,Eds.,pp.173-180,Elsevier Science Publishers (1984).
[2] 佐藤理史:MTB2:実例に基づく翻訳における複数翻訳例の組合せ利用, 人工知能学会誌,Vol.6,No.6,pp.75-85 (1991).
[3] 坂和正敏:ファジィ理論の基礎と応用,pp.30-43, 森北出版 (1989).
[4]Thomas H,et al:Introduction to Algorithms,pp.314-320,MIT press (1991).
[5] 安達久博, 下山豪彦:多機種間のマニュアル既要検索のためのファジー検索拡張シェル freshman, 情報処理学会第 49 回全国大会,3K-5 (1994).

[4]本システムは多機種間のマニュアル検索シェル freshman[5] に搭載される翻訳エンジン jfreshman としての位置付けである