日英比較表現のパターン化と意味的対応関係

斎藤健太郎　池原悟　村上仁一
鳥取大学大学院工学研究科
{ksaito,ikehara,murakami}@ike.tottori-u.ac.jp

1. はじめに

自然言語処理の技術的発展を図る上で意味解析の重要性が指摘されて意味のまとまりとなる原言語の表現をパターンとし、目的言語の表現と対応づける方法 [1] が提案されている。しかしこの手法では、日本語表現と英語表現の対応が単一的であり、文脈等を考慮したとき、対応している表現の中に適切な訳語表現がない場合がある。柔軟で適切な翻訳を実現する方法の1つとして、日本語と英語の表現を意味的特徴に着目して意味的類系ごとにまとめ、対応づける方法 [2] が考えられるが、意味類系を用いた手法では、訳文に用いる表現の候補が複数となるため、訳語表現を一意に決定できないという問題点がある。

そこで本研究は比較文型を例としてとりあげ、意味類系を用いた手法によってどの程度訳文に用いる表現の曖昧性が増加するかを調べ、複数の訳語表現の候補の中から、最も適切な訳語表現を一意に決定する手法を検討する。

2. 従来の研究と意味類系の必要性

2.1 研究の背景

従来、機械翻訳では、文を単語に分解して、単語単位に訳語を決定したのち、単語間の文法的関係に基づいてそれらを組み合わせ、訳語を合成してきた。しかし、現実の言語表現では文を単語に分解する過程で、文全体の意味が失われ、元の意味が再現されなくなることが多い。従来の機械翻訳システムでもこの問題を解決するために、連語や慣用表現などを単語に分解せずに、翻訳する方法など様々な方法が試みられている。しかし、この方法は逆に柔軟性に欠けるため、意味的に適切な翻訳結果が得られない場合が問題となる。

この問題を解決する方法の1つとして、意味的に表現を分類し対応づけることが挙げられる。すなわち、表現の意味的特徴に着目した意味類系を定義し、日本語、英語の表現を意味類系グループとしてまとめることである。この方法では、両言語間での表現パターン選択の自由度が増し、適切な対応関係が得られると期待される。

2.2 意味類系を用いたパターンの対応づけ

意味類系を用いた新しい方法の例を従来の方法の例と対比して示す。

見出し語動詞	日本語文型パターン	品詞の意味属性	英語文型パターン
入れる	N1がN2にN3を入れる	N1＜主体＞ N2＜施設＞ N3＜設備、機械＞	N1 install N3 in N2.
	N1がN2にN3を入れる	N1＜人＞ N2＜飲物＞ N3＜人＞	N1 make N2 for N3.

図 1: 日本語語彙体系に定義された文型パターン対の例

図1は日本語語彙体系 [1] に示されている表現対応づけの例である。日本語語彙体系では日本語表現は動詞と名詞の組合わせで文型パターンを登録しているが、表現が1対1でしか対応していないので、表現の対応が単一的であり、文脈等を考慮したとき対応している表現の中に意味的に適切な訳語表現がない場合がある。

これに対して、図2は意味類系を用いて日本語と英語の表現を対応づける方法の例である。意味類系ごとの対応表を用いることで、表現は複数対複数で対応される。複数対複数の対応になることによって、訳表現選択の自由度が増し、複数の訳表現の候補の中から意味的に最も適切な訳表現を決定できる可能性がある。

しかし、1つの原言語表現に複数の目的言語表現が対応するため、適切な目的言語表現を決定するための

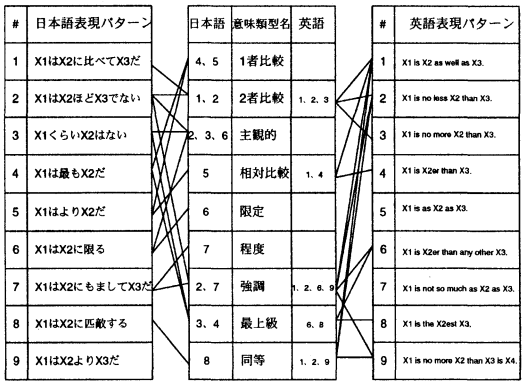

図 2: 意味類系を用いた日英表現パターンの対応例

仕組みが必要となる。

そこで、本研究ではこの問題に対するてがかりを得るため、比較文型に的をしぼって、意味類系を用いた時の日本語と英語の表現の対応関係を表にし、1つの表現パターンに対応している訳表現の候補の平均数や表現パターンが1対1で対応している割合を調べる。またその結果に基づき、複数の訳表現の候補の中から、最も適切な訳表現を一意に決定する手法について検討する。

3. 日英表現パターンの意味的対応関係の検討

3.1 用語の定義

本論文で用いる用語を以下のように定義する。

1. 比較文型…比較の意味 (日本語において) をもつ文。英語における as-as や日本語における" より" 等の単語によって構成される。

2. 意味キーワード…文の意味的な特徴を表す単語。文を意味的に検索する時の手がかりとなる。

   (例)as-as 構文における"as",「よりーだ」構文における" より"、" だ"

3. 表現パターン…文において意味キーワード以外の単語を変数においた意味を失わない範囲の最小単位

   (例)$N1$ is as K as $N2$.,「$N1$ は $N2$ より F で V する」

4. 訳表現…訳文に用いる表現、表現パターンのこと

   (例)$N1$ is as K as $N2$. に対する「$N1$ は $N2$ と同じくらい K だ」

5. 意味類系…表現パターンを意味ごとに分類する時に用いる分類器。意味的なパターンを示す。

   (例)「同級」「比較級」「最上級」

3.2 対応表の作成手順

本研究では図3に示した手順により、日英比較表現の対応関係を表にまとめる。

なお、用例収集では高校英語の教材を利用する。日本語の比較表現より英語の比較表現の方が参考書等から収集が容易で、また参考書にある表現は簡潔でわかりやすいためである。

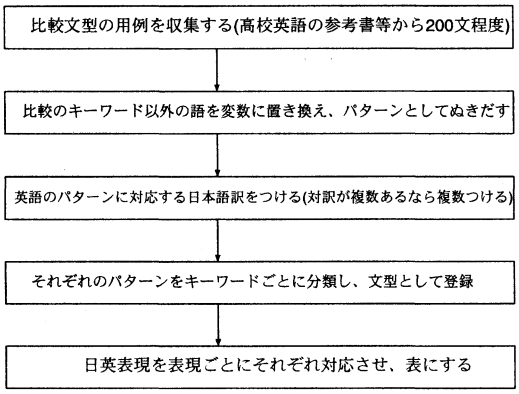

図 3: 日英対応表の作成手順

3.3 表現パターン対応表

図3の手順によって得られた日英表現パターンの対応関係を表1と表2に示す。表中の「番」とはそれぞれの表現パターンにつけた番号で、「対応番号」はその表現パターンに対応する訳表現候補の番号を表す。

表 1：英語から日本語への表現パターン対応表

番	英語表現パターン	訳日番
1	$N1$ is as K as $N2$.	1A.2A.3A…
2	$N1$ V ($N2$) as F as $N3$.	1B.2B.3B
3	$N1$ V ($N2$) as K $N2$ as $N3$.	…
4	$N1$ is not as $K1$ as $N2$.	5A.6A.7A
4a	$N2$ is not as $K1$ as $N1$.	…
4b	$N2$ is not as good as $N1$.	…
5	$N1$ not V ($N2$) as $F1$ as $N3$.	5B.6B…
5a	$N3$ V $N2$ not as $F1$ as $N1$.	…
6	$N1$ V ($N2$) as F as possible.	7A

表 2：日本語から英語への表現パターン対応表

番	日本語表現パターン	訳英番
1A	「$N1$ は $N2$ と同じくらい K だ」	1…
1B	「$N1$ は $N3$ と同じくらい F で V する」	2…
2A	「$N1$ は $N2$ のように K だ」	1…
2B	「$N1$ は $N3$ のように F で V する」	2
3A	「$N1$ と $N2$ を比べると同じだ」	1…
4A	「$N1$ は K において $N2$ に匹敵する」	1…
4B	「$N1$ は K においてに C (M) 匹敵する」	…
5A	「$N1$ は $N2$ ほどは $K1$ でない」	4…
5B	「$N1$ は $N3$ ほどは FV しない」	5…
6A	「$N2$ は $N1$ より $K1$ だ」	4…
6B	「$N3$ は $N1$ より FV する」	5…
7A	「$N1$ は可能な限り F で V する」	6

(1) 複数の訳表現候補に対応する例

表 1, 表 2 を利用することで、以下の例文のように表現パターンを複数の訳表現の候補として選択することができる。

(例 1) 英パターン 1：「$N1$ is as K as $N2$.」

例 1 に対応する表現パターンを表 1 から探し、対応している日本語の訳表現候補を表 3 に示す。

表 3:例 1 の表現に対応する訳表現候補

1A	「$N1$ は $N2$ と同じくらい K だ」
2A	「$N1$ は $N2$ のように K だ」
3A	「$N1$ と $N2$ を比べると同じだ」
4A	「$N1$ は K において $N2$ に匹敵する」

この例では英語パターン 1 に 4 つの日本語パターンが対応しており、表現の意味的な多義は 4 である。

(2) 表現が 1 対 1 に対応する例

次に訳表現が一意に決定できる例を以下に示す。

(例 2) 日パターン 7A：「$N1$ はできるだけ F で V する」

表 2 から例 2 の訳表現候補を探すと、英語パターン 6：「$N1$ V ($N2$) as F as possible.」のみ対応しており、訳表現は一意に決定する。

4 検討結果と考察

ここで表 1、2 より 1 つの表現パターンに対応する訳表現パターンの数を調べると、図 4、5 の結果を得る。また図 4、5 から表 4 の結果を得る。

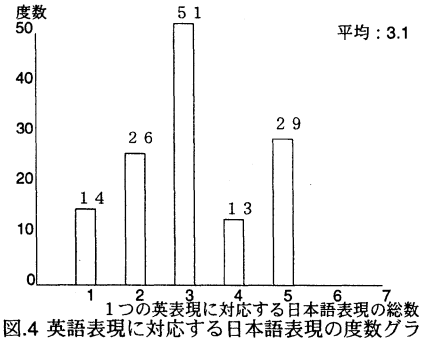

図.4 英語表現に対応する日本語表現の度数グラフ

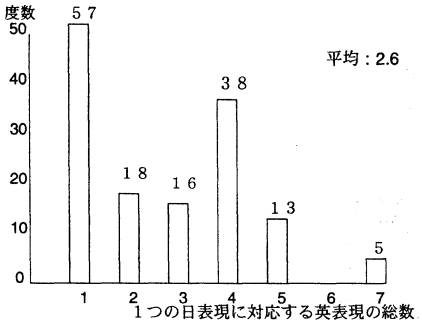

図5.日本語表現に対応する英語表現の度数グラフ

表 4:表現の対応に関する検討結果

	英 (から日)	日 (から英)
全表現数	139	148
1 表現の平均対応表現数	3.1	2.6
表現が 1:1 対応の割合	14/139	57/148

4.1 訳表現候補の平均多義数

表 4 の平均対応表現数から、英語と日本語の表現パターンは 3.1:2.6 の割合で対応していることがわかる。また表現パターン同士が 1 対 1 で対応する割合において、日本語から英語に対応させる方が高いことから、平均的にみて、日本語表現から英語表現に対応させる方が訳表現を一意に決定しやすいと言える。

4.2 意味類系ごとの特徴

比較文における代表的な意味類系 (同級、比較、最上) において、1 つの表現パターンに対応する訳表現候補の最大数を表 5 に示す。

表 5:意味類系ごとの訳表現候補の最大数

	同級	比較	最上
英から日	4	4	5
日から英	4	4	7

各表現パターンについての特徴は以下の通りである。

- 同級…同級以外の意味キーワードを用いた表現はほとんどない
- 比較級…同級の意味類系を表す意味キーワードを用いた訳表現でも表すことが可能
- 最上級…1つの表現に対応している訳表現の候補数が多い。同級、比較級の意味をもつ意味キーワードを用いた訳表現でも表すことができる

表5より、訳表現の候補の数が多い表現パターンが最上級の意味類系に属しており、そのことから最上級に属した表現パターンは多義性が高いことがわかる。最上級の表現で、対応する訳表現の多い例を表6に示す。

表 6:訳表現の候補の数が多い表現パターンの例

表現パターン	度数
No other $N1$ is as K as $N2$.	5
$N1$ is Ker than any other $N2$.	5
$N1$ は $N2$ では一番 K だ	7
$N1$ ほど K な $N2$ はない	7

4.3 訳表現の多義性の原因

表現パターンとそのパターンに使用される意味キーワードの関係について考える。ここで下記のパターンをとりあげると、この例3に対応する表現パターンの訳表現の候補は表7の3つのパターンである。

(例 3)$N1$ is Ker than any other $N2$.

表 7:例 3 の表現パターンの訳表現の候補

$N1$ はもっとも K な $N2$ だ
$N1$ は他のどの $N2$ よりも K だ
$N1$ は他のどの $N2$ に負けないくらい K だ

例3の表現パターンは最上級を表す意味類系に属している。しかし上の表現パターンに対する訳表現の候補は、最上級(「もっとも」最上級を表す意味キーワード)だけでなく(「より」比較級を表す意味キーワード)や同級(「くらい」同級を表す意味キーワード)を表す意味キーワードを用いて表すことが可能である。このことから、訳表現の多義性が発生する原因の一つに一つの意味キーワードが複数の意味類系にまたがって属していることがあげられる。この傾向は最上級の表現において顕著である。

4.4 訳表現の決定について

表7から一つの意味キーワードが複数の意味類系を表すことが、多義性発生の原因の一つであることがわかった。そこで意味キーワードが属している意味類系を調べると、最も適切な訳表現を一意に決定するための方法として意味キーワードに以下のようなパラメータを設けることが考えられる。

- 強調度…表現パターンが意味的に強調している度合を示す

  (例)「負けないくらい」 強調度:中,any other・「他のどの」 強調度:高,「最も」 強調度:低

- 比較度…表現パターンが意味的に比較している度合を示す。意味類系ごとに、意味的に同級に近付くほど低く、最上級に近付くほど高くなる

  (例) 「負けないくらい」 比較度:低,any other・「他のどの」 比較度:中,「最も」 比較度:高

例えば、例3の表現はは強調度:高, 比較度:中であるので、表5の3つの日本語訳表現候補の中から「$N1$ は他のどの $N2$ よりも K だ」(強調度:高, 比較度:中)が適していると考えられる。

5. おわりに

本研究は比較文型に分野を限定して、表現パターンに訳語表現を対応させた表を作成し、意味類系を利用した対応がどの程度曖昧性させるか調べた。またその曖昧性を解消するため、適切な訳語に用いる表現を決定する手法について検討した。

その結果、英語と日本語の表現の対応する比率は3.1:2.6で、日本語表現から英語表現に対応させる方が多いことがわかった。1つの意味類系に属した表現が複数のキーワードによって表わされることが多い点に着目すると、キーワードにパラメータを設けることで、最も適切な訳表現を一意に決定できる可能性がある。

今後このパラメータの詳細を検討すると共に、実験的にこの方法のは、本手法の精度を調べるため、選択された訳語表現の候補の中から適切な訳が含まれている割合を調べる必要がある。

参考文献

[1] 池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林:日本語語彙体系, 岩波書店 (1997)

[2] 池原 悟:意味類系に着目した日英意味辞書の研究企画, 辞書プロジェクト会議 (2000/4/1-2)