活用語を含む助詞的定型表現の分析

7B－2

辻孝子，安藤一秋，獅々堀正幹，青江順一
徳島大学工学部

1．　まえがき

複数の形態素列が一つの意味として解釈できる日本語の定型表現は，機械翻訳における正しい訳語の決定[1]に有用であり，また動詞を主体とする活用語を含む定型表現をまとめて処理することは，活用語に関係する統語や格構造解析を省略できるので，解析の効率化の点でも有効である[2]．特に，”助詞＋動詞（活用語）＋付属語”からなる形態素列が助詞相当語句に解釈できる表現（関係表現）は出現頻度も高く，その抽出方法が研究されている[2]．しかしながら，上記の有用性以外にも，文の短縮処理[3]，文書校正[4]（曖昧点の指摘）への応用を考慮した場合，更に進んだ分析が必要となる．従って，本報告では，活用語を含む定型表現で助詞相当語句に変換できる規則を定義し，その利用方法を提案する．

2．　文短縮に対する助詞的定型表現の利用

文書の記述できる範囲があらかじめ決まっている場合や，図表を適切に配置したい場合は，文の短縮を余儀なくされる場合が多々ある．執筆者や校正者の負担を軽減するためにも，コンピュータ支援が必要である．また，形態素列の短縮と削除を繰り返して，文章の骨格（要約）を抽出することに発展させることも可能である．これに対して，津田ら[3]は形態素列の置換規則を利用した文書の短縮手法を提案しているが，ここで議論する定型表現は考慮していないので，以下で提案する規則は，津田らの短縮規則を補足するものとなる．以下に津田らの短縮方法の１行分の短縮処理の概略を示す．なお，形態素Mに対する置換規則RULE(M)は，Mを置換するべき形態素を意味し，LENGTH(M)は形態素Mの長さを表すものとする．

【短縮処理の概要】

ステップ１：最後尾の行の文字数LAST(P)が最も少ない段落Pを見つける．

ステップ２：段落Pで，置換規則RULE(M)が適応可能な形態素Mを見つけ，RULE(M)で置換する．

ステップ３：LAST(P)から短縮分LENGTH(M)-LENGTH(RULE(M))を引いて，LAST(P)が零または負になれば終了し，そうでなければステップ２へ戻る．

たとえば，接続詞”しかしながら”をMとするとき，RULE(M)=”しかし”が置換すべき形態素となり，この置換で３文字分短縮されることになる．

津田らの短縮規則は，意味変化がない形態素置換を対象としているが，定型表現による短縮処理では，短縮の前後で文の解釈が変化することも生じる．次の文例で，意味変化を見てみよう．尚，この例では，下線で示す形態素列の置換候補を次の（）内に記述する．

(1)　定型表現<u>に関する</u>（の）処理を提案する
(2)　木<u>で作った</u>（の）椅子に座る
(3)　服<u>に付着している</u>（の）埃をはらう
(4)　法隆寺<u>に続いている</u>（への）道を進む
(5)　外国<u>に留学している</u>（の）友人を紹介する

文例(1)の定型表現「に関する」の助詞「の」への短縮は，文解釈に影響を与えないと思われるが，文例(5)の置換処理では，「留学している」が失われるので，元の文の意味を正しく解釈するのは，文脈を考慮しないと難しくなる．また，文例(2,3,4)の短縮処理で失われた意味は，定型表現に関係する名詞と文の主動詞から推論可能と思われる．

短縮処理でのステップ２で検出された短縮規則の候補の実行判断は，利用者にゆだねられるので，短縮処理が可能な全ての規則を優先順なしに検出してもよいが，利用者にとっては，意味変化の少ない短縮規則から確認できた方が能率的である．従って，短縮処理に利用する定型表現と助詞を意味変化の度合いで分類し，優先順位を付けておくことは効率的である．本研究では，短縮処理による意味変化についての実験を行い，意味変化の度合いについて分析してみた．

3．意味変化による定型表現の分類

3．1　定型表現による短縮規則

次の定型表現を短縮規則として，分類した．

K：活用語を含まない定型表現（格助詞相当語句）
　事故<u>が原因で</u>＝＞<u>で，により（省略も可）</u>
J：助述表現（活用語の付属語相当語句）
　記述<u>したものである</u>＝＞記述<u>した</u>.
R：活用語を含まない定型表現（K以外）
　しかしながら＝＞しかし
S：活用語を含むが，その意味が弱い慣用的な表現
　外国文化<u>に関する</u>研究＝＞外国文化<u>の</u>研究
T：活用語を含む定型表現（S以外）
　外国<u>に住んでいる</u>友人＝＞外国<u>の</u>友人
X：Tの定型表現で”こと”が付属する場合

Analysis of Frozen Patterns to Act as Propositional Patterns, by Takako Tsuji, Kazuaki Ando, Masami Shishibori and Jun-ichi Aoe, The University of Tokushima

Ｘ１：活用語がサ変動詞
　法案を提出することを＝＞法案の提出を
Ｘ２：活用語が一般動詞
　法案を出すことを＝＞法案の提出を
Ｘ３：活用語が形容詞
　画像が美しいことを＝＞画像の美しさを
Ｙ：”こと” 以外が付属する場合
Ｙ１：付属語が変化なし
　調査するとともに＝＞調査とともに
Ｙ２：語句を補う場合
　領域を設定するための＝＞領域の設定のための
　　[の を補う必要がある]
Ｙ３：語句を変更する
　日程を延期するので＝＞ 日程の延期のため
Ｚ：疑問詞節
　いつ大会は開催されるのか＝＞大会の開催予定
Ｂ：複合語への短縮［規則は宮崎の論文[5] 参照 ］

なお，Ｋ，Ｊは，首藤ら[6]による定型表現であり，置換処理に対して意味変化がほとんどないと考えられるものである．

３．２　意味変化の実験

ＥＤＲの日本語コーパスに対して，２０人の被験者に対して意味変化の実験を行った．その結果を図１に示す．なお，図中のＡ，Ｂ，Ｃは次の意味解釈を表す．

Ａ・・・得に変化なし
Ｂ・・・多少変化する
Ｃ・・・ かなり変化する

図１より，T分類以外の定型表現は，Ｃの解答がほとんどなく，置換処理しても意味に影響はないと思われるので，２章のステップ２で，適応の優先順位を考える必要はない．

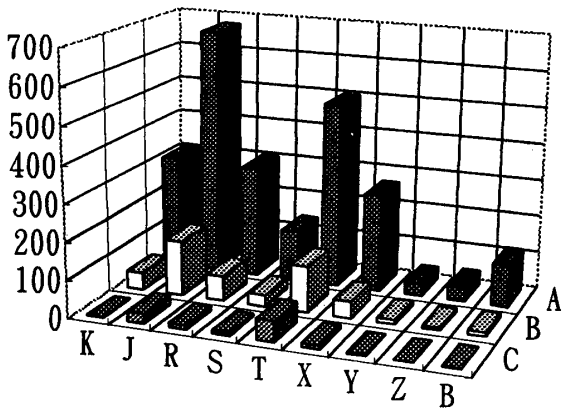

図１　意味変化の実験結果

３．３ 結果と考察

Ｔ分類のＣの解答が多いことについて検討する．

Ｔ分類では，N1+助詞＋動詞＋N2という形の下線部分を「の」，「への」に置き換えた．以下は，N1とN2 の意味素性をもとに，Ｔ分類の中でCの解答が多かった順に示したものである．

	N1		N2
1．	人間	—	具象物
2．	場所	—	人間
3．	目的	—	手段
4．	材料	—	制作物
5．	衣服類	—	人間
6．	組織	—	人間
7．	人間	—	贈り物
8．	身分・職業	—	人間

１．の，人間—具象物に対して，”母の絵”の場合，”母”と”絵”との関係は購買(母が買った絵)，所有（母が持っている絵），作成（母が描いた絵）など様々な意味が推測されうる[1]．この意味の決定にはN1やN2の意味素性，または文脈情報により，決定されるので[2]，形態素列の局所的情報を利用する文書短縮では，この分類による規則の優先順位は，低くすることが望ましい．

4．むすび

以上，本報告では，文書短縮の立場から活用語を含む助詞的定型表現の分析を行い，意味変化の度合いによる分類を行った．３．３節で示した優先順位の低い名詞の組が”の”で連結されているとき，曖昧性が多いわけであるから，文章推敲[4]では，逆に規則適用の優先順位が高くすることで，この分類は，有用になる．

今後，３．１節のＸ，Ｙ，Ｚの規則のように格構造が変化する場合の処理についても分析を進めたい．

参　考　文　献

(1)鳥津，内藤，野村：“助詞「の」が結ぶ名詞の意味関係のsubcategorization” 情処学NL研報,53-1 (1986)
(2)新納，井佐原：“疑似Nグラムを用いた助詞的定型表現の自動抽出” 情処学 Vol.36(1995-1)
(3)津田，中村，青江：“形態素置換による文書短縮法” 信学論 (D-Ⅱ), J75-D-Ⅱ.No3.pp.619-627 (1992-3)
(4)林，菊井：“日本文推敲支援システムにおける書き換え支援機能の実現方式” 情処学 Vol.32 (1991-8)
(5)宮崎：“係り受け解析を用いた複合語の自動分割法” 情処学 Vol.25 (1984-11)
(6)首藤：“文節構造モデルによる日本語の機械処理に関する研究” 福岡大学研報45自然科学編,6,pp.88-119 (1980-3)