概要

機械翻訳においての問題の一つとして複合動詞の翻訳処理がある。従来、複合動詞は訳語を個別にあらかじめ辞書に登録する手法がとられている。しかし、複合動詞の数は豊富で全てをあらかじめ辞書に登録するのは難しく、また、未知の複合動詞が出現した場合に新たに辞書登録する作業が必要となる。

そこで、本稿では辞書に登録する複合動詞の削減と未知の複合動詞への対応を実現するために複合動詞の構成要素から翻訳形式を決定するために構成要素の分類をした。この分類により複合動詞の構成要素である前方動詞は5段18種、後方動詞は2段24種に分類された。また、その分類を組合せ複合動詞の構成要素の分類表を作成した。そして、分類に使用した標本に付与されている訳語を参照し、分類表より導出できる翻訳形式を20個決定した。

次に分類表と分類表から導出される翻訳形式についての評価実験をした。一つ目の実験として、機能試験文集中の複合動詞を対象として翻訳形式と一致する割合を求め、翻訳形式の精度を評価した。二つ目の実験として、新聞記事中の複合動詞を対象として分類作成に使用した標本と同じ分類に属する割合を求め、一般文書への有効性を検証した。評価実験の結果として、翻訳形式と一致したのは対象の82%で、同じ分類に属したのは対象の70%であった。

以上により、本稿で手順を示し作成した分類は一般の文書にもある程度有効であることがわかり、分類より導出される翻訳形式の精度も高いことがわかった。

目 次

図目次

表目次

1　はじめに

機械翻訳においての問題点の一つとして複合動詞の翻訳処理があげられる。従来、複合語単位に訳語を辞書にあらかじめ登録する方法がとられている。その一つとして、最近、翻訳辞書の網羅性を向上させるために複合動詞とそれに対する日英対訳用例文の収集が進められている [?]。しかし、複合動詞の数は多いためあらかじめすべての複合動詞を辞書登録するのは困難であり、この手法では未知の複合動詞が出現した場合、その複合動詞の訳語の辞書への登録とそれに対する対訳用例文の収集が必要となる。

また、言語学的には複合動詞を構成要素で細かく分類している [?] ものもある。しかし、この分類は日本語側だけしか考慮しておらず、さらに、抽象的な概念を用いた分類も多い。これらの理由により、言語学的な複合動詞の分類そのままではほとんど機械翻訳には利用できない。そこで、日本語側から複合動詞を分類し、その分類から翻訳形式が決定できるようにすれば翻訳形式別に辞書登録することより登録数が削減でき、また、未知の複合動詞が出現しても構成要素を解析することにより翻訳形式の判別が可能となる。

本研究では複合動詞の構成要素から全体の翻訳形式を決定するための構成要素の分類を試みる。また、分類した構成要素を組み合わせて複合動詞の分類表を作成し、その分類表から翻訳形式を導出できるようにする。そして、決定した翻訳形式の精度を評価するために実際に複合動詞を分類し、実際の訳語と一致するかどうかを確認し、さらに、新聞記事を使って分類表との対応率を調べ、一般文書への有効性も検証する。

以下、第 2 章では複合動詞の形態について、第 3 章では分類の方法及び分類表作成について、第 4 章では分類表および翻訳形式の評価実験について、第 5 章で考察、第 6 章で今回の結論を述べる。

2　複合動詞とは

2.1　複合動詞の定義

「最小二つの実質的形態素が結合して、新しい文法的機能と意味を持つ大きな単位を形成するとき、そのまとまりを複合語という。そして、実質的形態素二つが共に動詞であるか、あるいは後部形態素が動詞であり、形成された複合語自体が一つの動詞としての文法的性質をもつまとまりを複合動詞と呼ぶ。」[?] と複合動詞は定義されている。つまり、2 個以上の単語からなる複合語であり、そして、複合語を構成する単語全てかまたは後部の単語が動詞であって全体として動詞の役割を果たすものが複合動詞である。

◎複合動詞を含む文章の例を以下に示す。
彼は急に走り出す。
空港から飛行機が飛んで行く。
今朝はまぶしい日差しで目覚めた。
私は街中を一日中ぶらついた。

2.2　複合動詞の形態

複合動詞の形態を表 1 に示す。表中の頻度は機能試験文集 [?] 中での頻度を示す。

表 1: 複合動詞の形態

	形態	例	頻度
(1)	動詞テ型 + 動詞	持っていく	72%
(2)	動詞原型 + 動詞	持ち上げる	26%
(3)	名詞 + 動詞	腰かける	1%
(4)	副詞/擬態語 + 動詞	べとつく	1%

（１）、（２）は意味的に分離できるものが多く構成要素の分類が可能であり、頻度においても複合動詞の大部分を占めているので本研究における分類の対象とする。しかし、（３）または（４）に当てはまる複合動詞は意味的に分離できないものが多く、それぞれの訳語を個々にあらかじめ辞書に登録する手法をとる必要がある。また、（１）や（２）に比べ（３）や（４）は頻度においてかなり少ないので（３）、（４）は本稿においての分類の対象としては取り扱わない。

ここで、本稿で扱う複合動詞を構成する単語はすべて動詞であるので、前部形態素（上記の動詞テ型、動詞原型にあたる部分）を前方動詞、後部形態素（上記の動詞にあたる部分）を後方動詞と以下定義する。

3　翻訳形式決定のための分類

複合動詞の翻訳形式を構成要素より決定する分類表を作成する。

3.1　構成要素分類ための標本

複合動詞の構成要素を分類するための標本として、本研究では機能試験文集 [?] 中 6183 文から人手により抽出した 675 個の複合動詞を使用する。（ここで、コーパスとして使用する機能試験文集とは日英機械翻訳の翻訳機能をテストするための表現を集めたもので、1 文ずつそれぞれに英訳が付与されている。）675 個の複合動詞の形態の内訳を表 2 に示す。

表 2: 標本の内訳

形式	個数
動詞テ型 + 動詞	498
動詞原型 + 動詞	177

標本として収集した機能試験文中の 675 個の複合動詞をリストにした。リストの一部を表??に示す。全リストは「参考資料 1：複合動詞リスト (1)、(2)、(3)」に示す。また、リストにした複合動詞を含む文の日本語文と英訳文は「参考資料 2：日本文—英訳文リスト (1)、(2)、(3)」に示す。(1) は「動詞原型 + 動詞」、(2) は「動詞テ型 + 動詞（「いる」を除く）」、(3) は「動詞テ型 + いる」の形式を集めたリストである。

3.2　分類方法

3.2.1　対訳による分別

標本に付与されている対訳を参照し、意味的に構成要素に分離できない複合動詞と分離できる複合動詞に分ける。本研究では翻訳形式を複合動詞の構成要素から決定するの

表 3: 複合動詞リスト（一部）

標準表記（現在/終止形)		標準表記	抽出した表記	原文参照番号
前方動詞	後方動詞			
出	会う	出会った	出会った	656/1616/4107
込み	合う	込み合う	込み合って	3152
助け	合う	助け合った	助け合った	5097
釣り	合う	釣り合う	つりあう	6947
落ち	合う	落ち合う	落ち合う	3134

で、構成要素に分離できない複合動詞は分類の対象とはしない。対訳による分別の結果を図??に示す。また、意味的に構成要素に分離できない複合動詞の例を表??に示す。

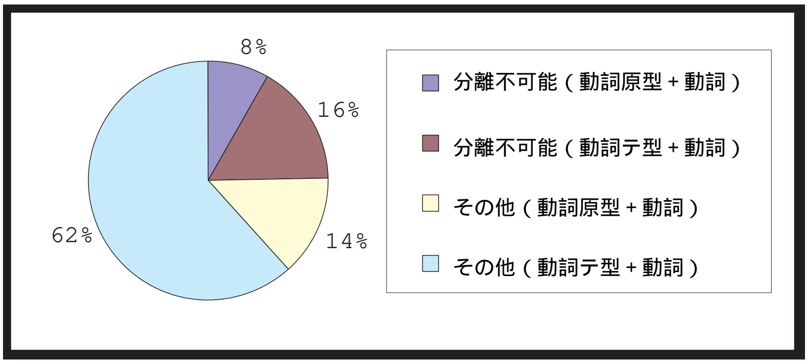

図 1: 意味的に分離できない複合動詞の割合

3.2.2 構成要素の分類

複合動詞の翻訳形式を複合動詞の構成要素から決定するために複合動詞の前方動詞、後方動詞に対してそれぞれを分類する。分類するにあたり翻訳形式がなるべく一意に決定できる深さまで分類をする。構成要素分類の手順を図??に示す。

構成要素分類の詳細を以下に示す。

（１）後方動詞の分類

後方動詞は機能動詞が多く種類も少ない。そこで、以下の 2 点で分類する。

1. 表記で分類 —「はじめる」、「いく」、「ある」等で分類
2. 動詞の持つ意味で分類 —「動作/現象の継続」、「現象/作用/状態の変化」、「空間的

表 4: 意味的に分離できない複合動詞の例

複合動詞	対訳	複合動詞	対訳
思い出す	remember	話し合う	talk
受け入れる	accept	引き合せる	introduce
繰り返す	repeat	申し込む	ask
見つける	find	やり遂げる	accomplish
出かける	go	言い張る	presist
出してくる	post	負けてしまう	lost
逃げ出す	run away	出会う	came across

図 2: 分類の手順

移動」等で分類
（ここで使用する動詞の意味の定義は[?]を参照した。）

分類した結果を木構造で表したものを図??に示す。

◎例 :「考えてみる」の「みる」の場合は以下のように分類する。
1. 表記 — みる、2. 動詞の意味 — 試みる
木構造上ではこのように分類される。

（２）前方動詞の分類
分類した後方動詞に対して接続する前方動詞を以下の1～5の順に分類する。

1. 活用形（形式）—「連用形」または「連用形＋て」で分類
2. 相状態 —「状態相」または「動作相」で分類
3. 種類 —「継続動詞」または「瞬間動詞」で分類
4. 意志性 —「意志動詞」または「無意志動詞」で分類
5. 動詞の持つ意味 —「行為」、「動作」、「位置/所有権の移動」等で分類

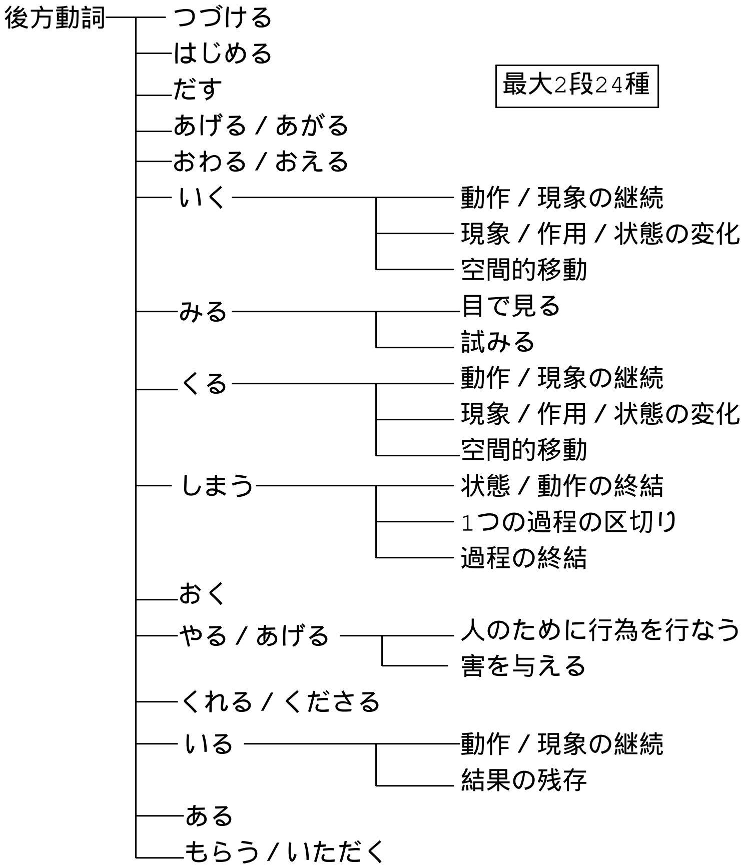
後方動詞
つづける
はじめる
だす
あげる / あがる
おわる / おえる
いく
動作 / 現象の継続
現象 / 作用 / 状態の変化
空間的移動
みる
目で見る
試みる
くる
動作 / 現象の継続
現象 / 作用 / 状態の変化
空間的移動
しまう
状態 / 動作の終結
1つの過程の区切り
過程の終結
おく
やる / あげる
人のために行為を行なう
害を与える
くれる / くださる
いる
動作 / 現象の継続
結果の残存
ある
もらう / いただく
最大2段24種

図 3: 後方動詞の分類木

（ここで使用する動詞の意味も後方動詞と同じく [?] を参照した。）

分類した結果を木構造で表したものを図??に示す。

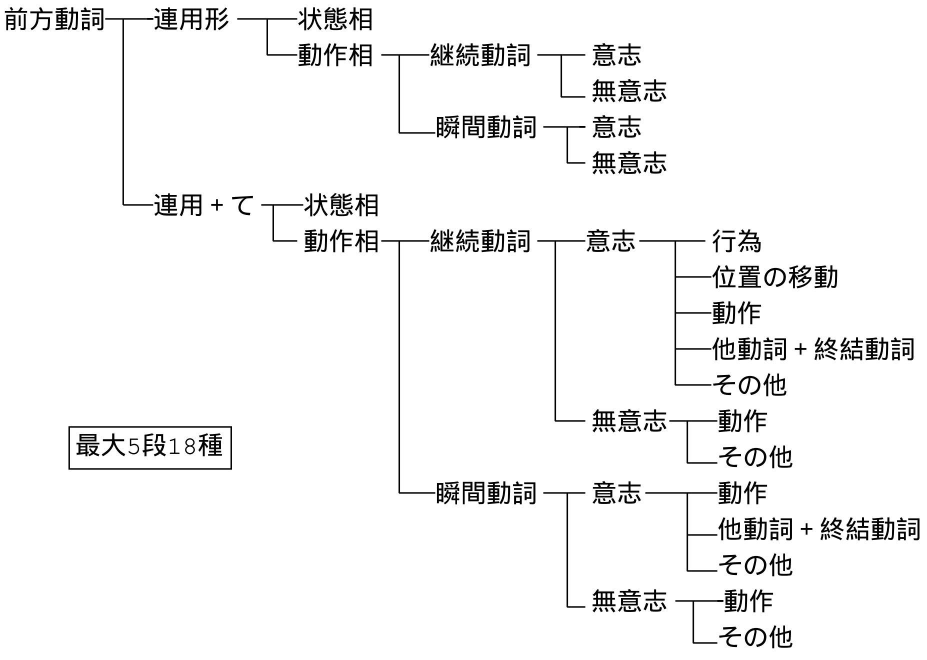

図 4: 前方動詞の分類木

◎例 :「走っていく」の「いく」の場合は以下のように分類する。

1. 活用形 — 連用＋て、2. 相状態 — 動作相、3. 種類 — 継続動詞、

4. 意志性 — 意志動詞、5. 動詞の意味 — 動作

木構造上ではこのように分類される。

3.3 分類表の作成

以上の分類の方法より、複合動詞の構成要素は前方動詞は 5 段 18 種、後方動詞は 2 段 24 種に分類された。その前方動詞、後方動詞の分類の組合せによりマトリクス状の分類表を作成した。作成した複合動詞の構成要素の分類表を表??に示す。表中の数値は標本の出現頻度を表す。

標本が出現した箇所それぞれの訳語の中で頻度の高い訳語の形式を分類表から導出できる翻訳形式と決定した。さらに、翻訳形式の形式ごとに$\#n$と番号をつけた。翻訳形

式番号を記した分類表を表??に、また、その翻訳形式番号から導出される翻訳形式一覧を表??に示す。

◎翻訳形式決定の例を以下に示す。

鳴らし (連用 + て、動作相、継続動詞、意志動詞、動作）+ 始める（発生）
→ 訳語：begin ringing
(連用 + て、動作相、継続動詞、意志動詞、動作）+（発生）に分類される複合動詞の訳語で頻度の高い訳語：begin doing（これを翻訳形式と決定）
→ #2 と番号をつける

表 5: 複合動詞意味構造組合せ表

表 6: 翻訳形式番号付き複合動詞意味構造組合せ表

表 7: 翻訳形式一覧表

4　実験

3 章で作成した複合動詞の構成要素の分類表について実験をし、翻訳形式の精度の評価と、一般文書への有効性を検討する。

4.1　実験方法

（１）機能試験文での翻訳形式の評価

機能試験文より人手により抽出した 395 個の複合動詞を対象として分類表に適用し、3 章で決定した翻訳形式と一致する割合を調べ、翻訳形式の精度を評価する。

（２）新聞記事での一般文書への有効性の検証

95 年毎日新聞経済欄 1 年文中より人手により抽出した複合動詞 1300 個を対象として、一つづつ人手により分類表に適用し、標本と同じ分類に属する複合動詞の割合を調べ、一般文書への有効性を検証する。

4.2　結果

実験の結果は以下の通りである。

（１）翻訳形式と一致した複合動詞数：323 個

機能試験文集中の複合動詞を対象として分類表に適用した結果、翻訳形式と一致した複合動詞の割合は 82%であった。翻訳形式と一致した複合動詞の分類表中での頻度を表??に示す。対象のどの複合動詞の訳語が翻訳形式と一致したかは「参考資料 3:翻訳形式一致複合動詞リスト (1)、(2)、(3)」に示す。(1)、(2)、(3) の区分は参考資料 1、2 と同じである。また、翻訳形式が一致しなかった複合動詞の実際の訳語は「参考資料 4:翻訳形式不一致複合動詞の訳語リスト」に示す。

（２）標本と同じ分類に属する複合動詞数：908 個

新聞記事中の複合動詞を対象として分類表に適用した結果、70%の複合動詞が標本と同じ分類に属した。新聞記事中の複合動詞を分類表に適用したときの出現頻度を表??に示す。なお、表中の黒枠は標本が出現した箇所を表す。

表 8: 翻訳形式一致数の内訳

表 9: 新聞記事中の複合動詞出現頻度

5　考察

5.1　翻訳形式の精度評価実験からの考察

実験では翻訳形式の精度は82%であったが、理由として本稿では分類するときに日本語側の複合動詞単体だけの情報で分類しており、英語側の情報や複合動詞が使用されている文中の動詞以外の情報などは用いていないことがあげられる。そこで、これらの情報も用いて分類を細分化すれば翻訳精度はさらに向上するのではないかと考えられる。

翻訳形式と一致した複合動詞、不一致であった複合動詞の例と不一致の原因を以下に示す。

○降り始める = start falling → 「start doing」 → 翻訳形式とは不一致
○鳴らし始める = begin ringing → 「begin doing」 → 翻訳形式と一致
上記の二つとも「~しはじめる」という複合動詞の訳語形式である。本研究では頻度の高いものを翻訳形式と決定したので「begin doing」のほうを翻訳形式として決定したが、「start doing」も翻訳形式になりえることは明らかである。しかし、「begin」になるか「start」になるかは複合動詞がとる主部の情報をもちいなければ判別することはできないと考えられる。そのため本研究で作成した分類表では対応できない。この点より、前方動詞、後方動詞の分類だけでなく第3の要素を用いて分類を細分化していかなければならない。

また、「~し始める」には「begin to do」という形式になる複合動詞もあるが「to do」か「doing」となるかは英語側に依存しており、あらかじめ動詞の翻訳辞書にどちらになるかを登録しておけばこの問題は解決できると考えられるので本研究ではどちらも翻訳形式として採用した。

5.2　一般文書への有効性検証実験からの考察

新聞記事に対しては70%の再現率であったが、これは、分類表作成に使用した標本の種類数が少なかったためと考えられるので、標本の規模を拡大すれば新聞記事のような一般文書に対しての再現率の向上が見込まれると考えられる。実験結果に関して注意すべき点として、対象全体では再現率は70%に留まったものの、個々の複合動詞的に見ると後方動詞が「いる」や「ある」である複合動詞のように再現率が100%のものもあれば、後方動詞が「つづける」や「いく」や「くる」である複合動詞のように再現率が全体の再現率より低いものもある点があげられる。この点より、今後、標本の規模を拡大

するにあたり本実験で特に再現率の低かった複合動詞に関して重点的に収集をし、種類数の不足を補う必要があると考えられる。

6　結論

本稿では辞書に登録する複合動詞の削減と未知の複合動詞への対応を実現するために複合動詞の構成要素から翻訳形式を決定するために構成要素の分類をした。また、その分類を組合せ複合動詞の構成要素の分類表を作成した。そして、分類に使用した標本に付与されている訳語を参照し、分類表より導出できる翻訳形式を決定した。

複合動詞の構成要素を分類した結果、前方動詞は5段18種に後方動詞は2段24種に分類された。また、その分類の組合せの分類表から導出できる翻訳形式を20個決定した。

評価実験として、機能試験文集から抽出した複合動詞を対象として、分類表から導出される翻訳形式と適合する複合動詞の割合を調べ、翻訳形式の精度を評価した。また、95年毎日新聞経済欄の複合動詞を対象として、実際に分類表に適用し標本と同じ分類に属する複合動詞の割合を調べ、一般文書への有効性を検証した。

評価実験の結果として、機能試験文集中の複合動詞の82%が分類表の翻訳形式と一致し、新聞記事中の複合動詞は70%が標本と同じ分類に適用された。

今後の課題としては、大量の標本を用いて翻訳形式の定義洩れを補完していく必要と、英語側の情報や複合動詞を含む文の主部の情報などを用いて分類を細分化する必要がある。

参考文献

[1] 白井、大山、武智、分部、相澤: 複合和語動詞に対する日英対訳用例文の収集について第57回情報処理学会全国大会,Vol2,P267-268,1998

[2] 新美和昭、山浦洋一、宇津野登久子: 複合動詞, 荒竹出版, 1993.

[3] 池原、白井、小倉: 言語表現体系の違いに着目した日英機械翻訳機能試験項目の構成人工知能学会誌,Vol9,No4,P569-579,1994