

概要

結合価文法による日英機械翻訳において、結合価パターンの表す用言と格要素の意味的關係が、動詞と名詞の訳語選択に有効であると考えられている。しかしその効果はまだ検証がなされていない。本研究では、動詞と名詞の訳語選択における、結合価文法の効果を定量的に検証する。評価対象は、IPAL 辞書に登録されている基本動詞と基本名詞に関する対訳(単文)とする。結合価文法を用いた翻訳システムに ALT-J/E を用いて翻訳し、その結果と対訳英文を比較したところ、基本動詞および基本名詞の正解率は、それぞれ 89% および 91%であった。この正解率はパターン決定の際に頻度情報が用いられ、人手で一意に決定できないものについても正解する場合があったと考えられる。しかしその結果も訳語選択の基本に結合価文法を用いているので、それも含めて結合価文法の訳語選択能力の評価とした。また失敗例の解析から、パターンの拡充などにより動詞について約 9-10%、名詞について約 3%の精度向上が見込まれると推測される。よって動詞 98%、名詞 94%が結合価文法を用いた訳語選択の限界と推定される。

目次

1	まえがき	1
2	結合価文法による訳語選択	2
2.1	結合価文法	2
2.2	一般名詞意味属性体系	3
2.3	結合価文法による訳語選択	4
2.3.1	動詞の訳語選択方法	4
2.3.2	名詞の訳語選択方法	5
3	評価実験	6
3.1	実験のねらい	6
3.2	実験システム	6
3.3	評価対象	8
3.4	評価方法	9
3.5	評価基準	9
4	実験結果	11
5	考察	12
5.1	動詞の訳語選択に関する考察	12
5.2	名詞の訳語選択に付いての考察	17
6	まとめ	23

目 次

1	一般名詞意味属性体系	3
2	パターン適用の手順	4
3	パターンによる名詞意味属性の絞り込み	5
4	人手によるパターン適用の問題	7
5	ALT-J/E における頻度情報によるパターン選択	7
6	IPAL 辞書の書式	8

表 目 次

1	結合価パターン対の例	2
2	動詞の訳し分け精度	11
3	名詞訳し分けの精度	11
4	パターンの数別の正解率	12
5	正しい動詞訳語が出せなかった原因	13
6	正しい名詞訳語が出せなかった原因	17
7	意味属性による訳し分けの可能性と実測値	21

1 まえがき

機械翻訳において訳語選択は、重要であると同時に大変困難な問題の一つである [1]。特に日本語と英語のように言語族の異なる言語間においては、対象の捉え方の違いによって訳語選択の問題をより困難にしていると考えられている。また日英機械翻訳における訳文品質の分析 [2] において、訳語選択の不適切さが訳文品質低下の原因の 40% を占め、構文解析の誤り (25%) を抜いて最大であると報告されている。訳語選択の精度を向上させるためには、文中で使用された単語の意味 (語義) を解析する必要がある。

これまでに動詞の意味解析の方法としては共起レベルと頻度をパラメータとした「尤度」、および辞書的記述に基づいた「制約」を用いる方法 [3] や単語の共起関係を用いた方法 [4] などが提案されているが、訳語選択において高い精度を出すに至っていない。また、名詞の意味解析の方法としては、談話解析で得られた場面情報により、英語名詞の多義を解消する方法 [5] や連体修飾句の係る名詞を対象に格の情報と意味属性を用いる方法 [6] が提案されている。最近では名詞辞書における多義構造の記述に関する研究 [7][8] もあるが対象は限定的である。

このような状況の下、多段翻訳方式が提案され、日英機械翻訳システム「ALT-J/E」が開発された。この翻訳方式では結合価文法 [9] に基づいた単文レベルの日英文型パターン対を用いて日本語から英語へ表現の変換を行なっている。ここで結合価文法はその構造から、動詞と名詞の訳し分けに有効と考えられるが、その効果は定量的には明らかにされていない。

そこで本研究では、結合価文法による動詞と名詞の訳語選択の有効性を定量的に検証するために、ALT-J/E の訳文評価実験を行ない、その効果を定量的に明らかにする。ここで基本的な動詞および名詞の訳し分けを評価するために、「計算機用日本語基本動詞、名詞辞書 IPAL」 [10] (以下、「IPAL 辞書」とする) に登録されている基本動詞と基本名詞を含む単文の対訳を対象とする。IPAL 辞書に登録されている語は日常的に用いられる語を網羅しており、また基本的な語ほど多義を持つという観点から本稿の趣旨に適切と考え用いるものとする。次に評価実験の結果から訳語選択を誤った文について、今後結合価文法で訳語選択できる語の種類と、訳し分けに必要な知識について検証を行ない、結合価文法の有効性の範囲と限界を明らかにする。

本稿では以下、第 2 章では結合価文法による訳語選択、第 3 章では評価実験、第 4 章では実験結果、第 5 章に考察、第 6 章でまとめを述べる。

2 結合価文法による訳語選択

2.1 結合価文法

結合価文法による翻訳では，結合価パターン(以下「パターン」とする)を用いて訳語を決定する．パターンは，用言を中心に意味的に必須の格要素(名詞+助詞)を記述したものである．この意味的關係の記述により用言と名詞の間に意味的な制約が生まれ，日本語解析で発生する意味上の多義が解消されることが考えられる．

本研究では，日本語語彙体系 [11] に記載されている「構文意味辞書」のパターンを用いる．パターンは表1のように，各見出し語の用法ごとに日本語文型パターンが登録されており，それぞれに英語の文型が1対1対応で登録されている．各動詞(用言)は用法ごとに，各見出し語につき1つ以上のパターンが登録されており，全体では14,000件のパターンが登録されている．表1に動詞『送る』を見出し語とするパターンを例示する．表1の例で変数『N』は体言が対応する．また，かっこ内は一般名詞意味属性(以下「意味属性」とする)であり，対応する体言を制約する．(意味属性『*』は制約なし)

表 1: 結合価パターン対の例

日本語文型	英語文型
N1(人)がN2(休暇)を送る	N1 spend N2
N1(人)がN2(生活)を送る	N1 live N2
N1(主体)がN2(使者・探偵 or 団体)をN3(主体 or 場所)に/へ/まで送る	N1 dispatch N2 to N3
N1(主体)がN2(*)をN3(主体 or 場所)に/へ/までN4(通信機器)で送る	N1 send to N2 N3 by N4
N1(人)がN2(人 or 乗物)をN3(場所 or 場)に/へ/まで送る	N1 see N2 to N3
N1(主体)がN2(主体)をN3(建造物 or 場所 or 場)で送る	N1 see N2 off at/in N3
N1(主体)が拍手をN2(*)に送る	N1 applaud N2
N1(主体)が声援をN2(主体)に送る	N1 cheer up N2
N1(主体)がN2(語 or 文字類)をN3(文・句・語 or 記号(言語))に送る	N1 add N2 to N3
N1(主体)がN2(抽象)をテレックスで送る	N1 telex N2
N1(主体)がN2(合図)をN3(主体)に送る	N1 give N2 to N3
N1(主体)がN2(主体)に賄賂を送る	N1 bribe N2

2.2 一般名詞意味属性体系

一般名詞意味属性は、名詞の意味的用法に着目してシソーラスとして体系化されている。本研究で使用する一般名詞意味属性体系は、最大12段の木構造であり、2,710の意味属性の分類があり、そして約40万語の名詞が登録されている。また、この木構造において上位の意味属性の性質を下位の意味属性が継承する。この特徴により、パターンにおける変数『 N 』への制約を簡潔に記述することができる。また、その制約によって『 N 』に対応する名詞の意味を絞り込むことができる。

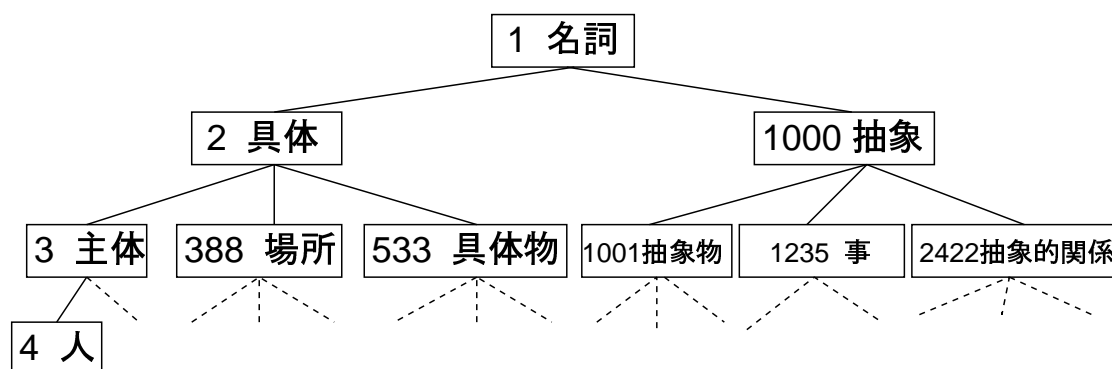


図 1: 一般名詞意味属性体系

2.3 結合価文法による訳語選択

結合価文法を実際に日本語文に適用して訳語を選択する方法を説明する。

2.3.1 動詞の訳語選択方法

まず対象とする文の用言から、その用言を見出し語とするパターンをパターン辞書から検索する。次に格要素の意味属性、格助詞などが最も多く一致するパターンを一つ選択する。パターンは日英のパターン対で登録されているため、パターンが決定することにより動詞の訳語が決定する。

図2で、例文「彼は友人を家まで送った」に対して実際にパターンを適用する方法を説明する。まず例文の用言が『送る』なので、パターン辞書から動詞『送る』を見出し語とするパターンを検索する(図2(a))。次に検索したパターンと例文を比較する(図2(b))。すると、格要素の数、意味属性、格助詞が最も適合する(3)のパターン『N1(主体)がN2(主体)をN3(場所)に/へ/まで送る』がこの例文に対するパターンとして選択される。そして、パターンが決まることにより『送る』の訳語は『see ... to』と選択される。

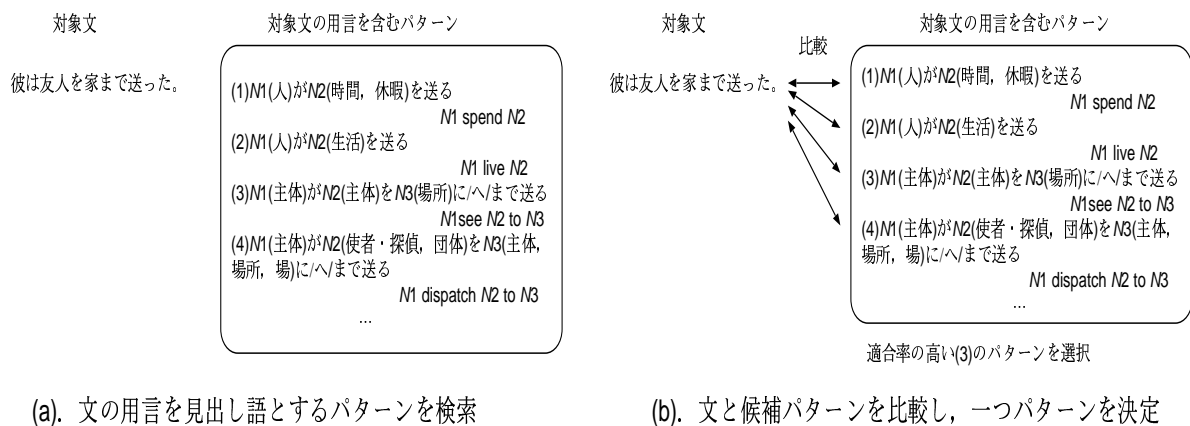


図 2: パターン適用の手順

2.3.2 名詞の訳語選択方法

次に日本語パターンの意味属性から、対応する名詞の訳語が絞り込まれる様子について説明する。パターンには『N1(場所)』といった体言 N に対する制約が記述されている。よって『N1(場所)』とあれば、 $N1$ は意味属性<場所>配下の意味属性に制限される。本研究で扱う名詞は、単語辞書において、名詞の見出し語、その名詞が持つ意味属性と対応する英訳語が記述されている。よって意味属性が制約されることにより訳語を決定することができると考えられる。

具体的な例を挙げて意味属性による名詞の訳語の絞り込みについて説明する。名詞『家』には前述した ALT-J/E の辞書において、3つの意味属性<家族>、<居住施設>、<家屋(本体)>が記述されている。そしてそれぞれ、意味属性<家族>には英訳語『home』が、<居住施設>、<家屋(本体)>には英訳語『house』が付与されている。ここで前項の例文「彼は友人を家まで送った」を考える。前述した通りこの例文にはパターン『N1(主体)がN2(主体)をN3(場所)に/へ/まで送る』が適用される。このパターンから名詞『家』は『N3(場所)』の意味属性に制限される。ここで名詞『家』の3つの意味属性のうち、この制約によって残るのは、『図3』に示すように<居住施設>だけとなる。よって名詞『家』の訳語は意味属性<居住施設>に付与されている『house』に決定される。

以上の手順より動詞と名詞の訳語が決定するので、結合価文法を用いることにより動詞と名詞の訳語選択に効果があると考えられている。

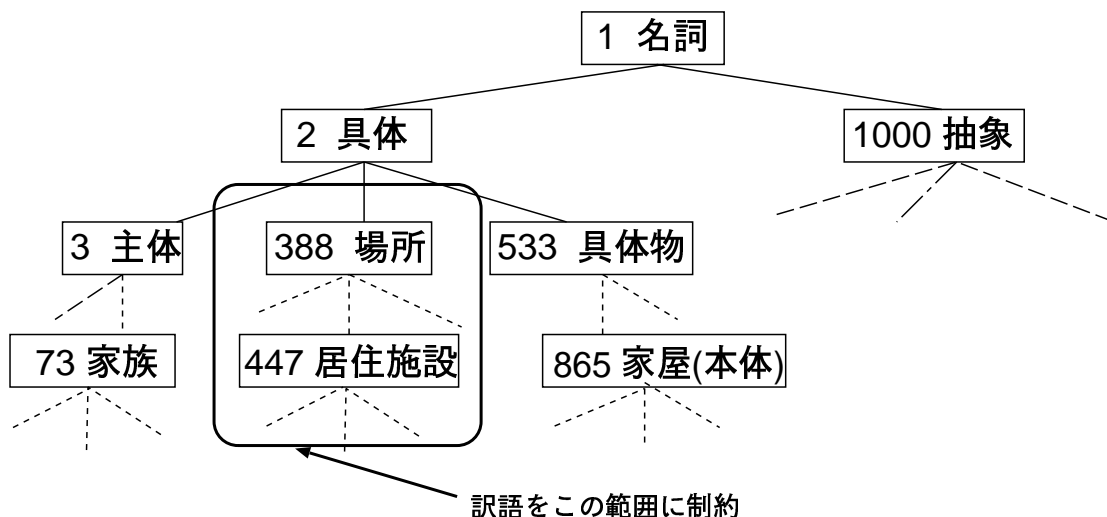


図 3: パターンによる名詞意味属性の絞り込み

3 評価実験

3.1 実験のねらい

2章で述べた結合価文法による，動詞と名詞の訳語選択の性能を調査するために評価実験を行なう．評価実験では，結合価文法によって導き出された英訳語と正解例の英訳語を比較して行なう．

3.2 実験システム

結合価文法を実際に人手で適用することは困難なため，翻訳ソフト「ALT-J/E」(以下ALTとする)を用いる．ALTは結合価文法を用いて訳語選択を行なう．よって本研究ではALTの結果から結合価文法の効果を検証する．実際に人手でパターンを適用すると，一意にパターンを決定できない場合が有る．図4の例「母が大根をおろした」では『大根をすりおろした』のか『大根を(市場などへ)卸した』のか判断ができない．こうした場合，人手でパターンを一意に決定することができない

一方ALTでは，人手で一意に判断できない場合についても頻度情報によって一意にパターンを決定する．ここで頻度情報とは各パターンの使用頻度などから各見出し語の各パターンに重みを付与したもののことである．格要素などの情報から一意にパターンを決定できない場合に，ALTではこの頻度情報によって一意に訳語を決定する．よって，ALTを用いることにより，人手で行なう場合に比べて精度が高くなる可能性があると考えられる．しかしその結果も訳語選択の基本に結合価文法を用いているので，それも含めて結合価文法の訳語選択能力の評価とする．

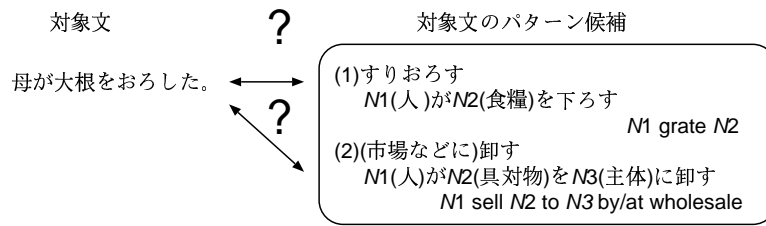


図 4: 人手によるパターン適用の問題

図 5: ALT-J/E における頻度情報によるパターン選択

3.3 評価対象

評価実験の対象は、IPAL 辞書に登録されている基本動詞および基本名詞を含む単文の日英対訳文とする。IPAL 辞書は日本語で頻繁に用いられる基本動詞 (861 語)、および基本名詞 (1,081 語) をそれぞれの用法ごとにまとめたものである。日常的に頻繁に用いられる基本的な単語ほど多義を持つ、という観点から IPAL 辞書は訳語選択実験の評価対象として適切であると考えられる。

IPAL 辞書は、まず見出し語があり、その見出し語ごとに用法が検討され、各用法ごとに用例が記載されている。例を図 6 に示す。図 6 では下線部 1 が見出し語を、下線部 2 が見出し語の意味を、下線部 3 が用例をそれぞれ示している。

しかしこの用例は日本語のみであるため、基本動詞の用例については、翻訳家に依頼して対訳を追加した。本研究ではこの対訳を正解例とする。動詞の訳語選択の評価実験においては、この対訳を用いる。名詞の評価実験においては、上記の対訳および機能試験文集 [12] から基本名詞を含む単文の用例をランダムで抽出したものをを用いる。評価実験は動詞に関する 5242 文の例文と、名詞に関する例文 1062 文の例文について行なう。

なお IPAL 辞書の日本語の用例は、ほとんどが単文であり、ALT の係り受け解析の誤りなどが起きにくいという点からも、訳語選択の精度調査に適していると考えている。例文全文のデータは巻末の付録に掲載する。

① 見出し語 ② 見出し語の意味 ③ 用例

①"あう", "0 0 1", "0 0 1", "0 1 0", "最重要動詞 1 2 3", ②"複数の物の断面がぴったりする。", "くつつく、一致する、合致する", "合い・組み・解けなど", "2. 1 5 5 0", "結合", "2 2 0 b", "動作 (動き), (抽象的) 関係, (状態) 変化", "合 (あ) う", "五段", "a w", "自", "あわせる", "N 1 ガ", ③"雄型と雌型がぴったりと 合った。", "O 1", "ガ", "PRO / ABS", "雄型と雌型", "鍵と鍵穴 / 断面と断面", "×", "×", "相互", "現在", "O", "テシマウ", "願望", "推量", "ナ (願望)", "2", "マス, ナイ", "N 1 は複数主体。"

"あう", "0 0 1", "0 0 2", "0 1 0", "最重要動詞 1 2 3", "複数の物がまとまって一つになる。", "合する、接する、くつつく", "合い・組み・解けなど", "2. 1 5 5 0", "結合", "2 2 0 b", "動作 (動き), (状態) 変化", "合 (あ) う", "五段", "a w", "自", "あわせる", "肌あい", "N 1 ガ", "二つの川が この地点で 合う。", "まぶたが 合う。", "O 1", "ガ", "NAT / PRO / PAR", "三つの川 / 国道と県道、鉄道 / 唇、まぶた", "×", "×", "相互", "現在", "O", "×", "願望", "推量", "×", "2", "マス, ナイ", "視線が〜。目 (※と目) が〜。", "N 1 が PAR の時、「あわせる」という自動詞あり。", "N 1 は複数主体。"

...

図 6: IPAL 辞書の書式

3.4 評価方法

評価はALTで機械翻訳した英訳文のうちの注目単語の英訳語と対訳の英訳語とを比較して行なう。評価は『 』 『 × 』の三段階とする。判断は評価者1名によって行なう。

3.5 評価基準

『 』 『 × 』の三段階の評価基準を以下に示す。ALTによる機械翻訳において、ALTの構文解析などが原因で例文が部分的に正しく翻訳されない場合がある。その場合、本研究の趣旨から対象とする語が正しく英訳されている場合は正解とする。評価基準は動詞、名詞のいずれに関する例文の場合も同じ基準で評価を行なう。

評価 〇 : 対象とする単語のALTの英訳語が、対訳例が用いている訳語と一致する場合。

例1 : 二つの川がこの地点で合う。

対訳1 : The two rivers join at this point.

ALT訳1 : Two rivers join in this point.

例2 : 彼は皆の注目を集めた。

対訳2 : He attracted the attention of everyone.

ALT訳2 : He attracted the attention of everybody.

評価 △ : 対象とする単語のALTの英訳語が、対訳例が用いている訳語と異なるが、意味的に正しい場合。

例3 : 彼は準備を急いだ。

対訳3 : He prepared quickly.

ALT訳3 : He hurried preparation.

例4 : 私たちは今日その問題を話し合った。

対訳4 : We discussed the matter today.

ALT訳4 : We talked about that problem today.

評価 × : 対象とする単語のALTの英訳語が間違っている場合。

例5 : 彼らは海底に沈んでいた船を陸に揚げた。

対訳 5 : They salvaged the sunken vessel on the bottom of the sea.

ALT 訳 5 : They deep-fried in land the ships that had sunk in the bottom of the sea.

4 実験結果

実験の結果をデフォルトの訳語による評価結果と共に表 2 , および表 3 に示す . デフォルト訳語とは , 各見出し語の訳語の中で最も良く使われる語のことで , PROCEED 和英辞書 [13] で先頭に表記されている語をデフォルト訳語とする . 表 2 に動詞の訳語選択の正解率を , 表 3 に名詞の訳語選択の正解率をそれぞれ示す . なお , 全文の評価結果は巻末の付録に記載する .

表 2: 動詞の訳し分け精度

評価	ALT-J/E	デフォルト訳語
	49% 2572 文	22% 1141 文
	40% 2081 文	33% 1740 文
×	11% 589 文	45% 2361 文
合計	5242 文	5242 文

表 3: 名詞訳し分けの精度

評価	ALT-J/E	デフォルト
	62% 658 文	58% 615 文
	29% 312 文	27% 289 文
×	9% 92 文	15% 158 文
合計	1062 文	1062 文

5 考察

5.1 動詞の訳語選択に関する考察

5.1.1 動詞の訳語選択への有効性

表2よりALTの正解率(評価 と の合計)がデフォルトの訳語に比べ34%高かった。このことから動詞の訳語選択において結合価文法の効果が有ったと言える。

ここで、各動詞に登録されているパターン数と正解率の関係について考える。多義の少ない語であれば、その訳語を使い分けるために多くのパターンに登録する必要はない。逆に多義の多い『ある』のような語に対してはパターンを多く登録する必要がある。よって訳語を細かく使い分けるにはパターンを追加する必要があるが、パターンが多いと選択が困難になるという問題を抱えることになる。

そこで、各動詞について登録されているパターンの数ごとに動詞を分類し、パターンの数による訳し分けの精度を調査した。結果を表4に示す。

その結果から、パターン数が0個の動詞、つまりパターンが未登録の動詞の誤り率(評価×)が22%であるのに対して、パターンに登録されている場合には約半分となっていることが分かる。このことから、結合価文法を用いることで動詞の訳語選択の精度を上げることができると言える。しかしパターンに登録されているものについても、パターンが多くなると文に対応したパターンの選択に失敗し、訳語選択に失敗するという傾向が見られ、今後、不足するパターンの追加と共にその選択方法も問題となると考えられる。

表 4: パターンの数別の正解率

パターン数	0 個	1 ~ 5 個	5 ~ 10 個	11 個以上
	50%	48%	56%	50%
	28%	43%	37%	37%
×	22%	9%	7%	13%

5.1.2 訳語誤選択の原因と解決方法

ALT が正しい訳語を選択できなかった原因を明らかにするために、評価が×となった例文 589 文中 122 文について調査した。結果を表 5 に示す。なお、各原因に該当する例文は巻末の付録に記載する。

表 5: 正しい動詞訳語が出せなかった原因

	翻訳失敗の原因	割合
1	パターンが登録されていない場合	21% 26 文
2	パターンの照合に失敗した場合	37% 45 文
3	慣用表現が用いられている場合	11% 13 文
4	形態素解析に失敗した場合	9% 11 文
5	係り受け解析に失敗した場合	17% 21 文
6	例文が複数の意味にとれる場合	5% 6 文
	合計	122 文

ALT では、動詞の訳語選択において表 5 に示す問題を含んでいる。将来、これらの問題が解決されれば、さらに結合価文法により正しい訳語を選択できるようになると考えられる。以下にそれぞれの問題の解決方法を考察する。

(1) パターンが登録されていない場合

対象の文に対するパターンが登録されていない場合、パターンを用いた訳語選択ができない。具体例を例 6 及び例 7 に示す。

例 6 : 学生が教授に教授の都合を電話で 伺った。

対訳 6 : The student phoned the professor and asked him when he would be free.

ALT 訳 6 : A student listened the professor's circumstances with a telephone to a professor.

例 7 : 活字がいくつか 落ちている。

対訳 7 : Some characters are missing.

ALT 訳 7 : A type is falling variously.

例 6 の文に対して、動詞『伺う』のパターンは「N1(人)がN2(人)のN3(陳述)を伺う N1 listen to N2」としか登録されていないため、『伺う』を用言として持つ文に対してはこのパターンしか適用されず、他の訳語を選択できない。この対策としては、不足

しているパターンを補うことがあげられる．具体的に例 6 の場合「 $N1$ (人)が $N2$ (人)に $N3$ (状態)を伺う $N1$ ask $N2$'s $N3$ 」をパターン辞書に追加することにより目的の訳語『ask』が選択できると推測される．同様に例 7 についてもパターン「 $N1$ (文字)が落ちる $N1$ miss」を追加することによる解決できると考えられる．

(2) パターンの照合に失敗した場合

一つの動詞に対して複数のパターンが登録されていると，日本語文に対するパターンの選択が問題になる．具体例を例 8 及び例 9 に示す．

例 8：彼は海外で夏休みを送った。

対訳 8：He spend his summer vacation abroad.

ALT 訳 8：He saw a summer vacation off at a foreign country.

例 8 において，動詞『送る』の場合 12 種類のパターンが日本語語彙体系に登録されている．この例文に対しては「 $N1$ (人)が $N2$ (時間，休暇)を送る $N1$ spend $N2$ 」のパターンを適用すれば，正しい訳語が導き出される．しかし実際には「 $N1$ (人)が $N2$ (主体)を $N3$ (場所，場，建造物)で送る $N1$ see $N2$ off at/in $N3$ 」が適用され，誤った訳語が選択されている．この問題は対象の文とパターンを照合させるアルゴリズムの問題であり，現在改善方法が研究されている [14]．

例 9：彼はフグに当たった。

対訳 9：He was poisoned by a puffer fish.

ALT 訳 9：He consulted フグ.

例 9 では名詞『フグ』が辞書登録されていないために翻訳されていない．これにより，パターン「 $N1$ (人)が $N2$ (魚介，食糧)に当たる $N1$ be poisoned by $N2$ 」の『 $N2$ (魚介，食糧)に』の部分に当たる名詞の意味属性を判断できない．よってパターンを照合する際の情報が不足し，パターンの判断を誤ったと考えられる．この問題については例 9 中の名詞『フグ』を辞書登録されている名詞『河豚』に書き換えて，ALT で再実験を行ったところ「He was poisoned by a globfish.」と，正しい訳語を得ることができた．よって例 9 のような場合には訳せなかった名詞を辞書登録することで解決できると考えられる．

(3) 例文に慣用表現が用いられている場合

慣用表現が用いられている場合の例を例 10 及び例 11 に示す．

例 1 0 : 彼は 話の腰を折った .

対訳 1 0 : He interrupted a person's speech .

ALT 訳 1 0 : He broke the waist of talk .

例 1 1 : 彼女は膝の上に 眼を落とした .

対訳 1 1 : Her eyes fell to her lap. .

ALT 訳 1 1 : She dropped an eye on the top of her knee.

慣用表現は言い回しが決まっているので、パターンの記述に意味属性を用いずに、表現に含まれる名詞の字面を記述することができる。具体的には例 1 0 に対しては、「N1(人)が話の腰を折る N1 interrupt a speech」のパターンを登録することで正しい訳語が選択できる。同様に例 1 1 についても「N1(人)が眼を落とす N1 N1's eyes fell」といったパターンを追加することで解決できる。

(4) 形態素解析に失敗した場合

(5) 係り受け解析に失敗した場合

(4) および (5) の場合は結合価文法を用いる以前の処理の問題である。これらの処理が正しく行われないと、結合価文法を使うことができない。本研究では結合価文法による訳語選択能力の調査が目的なので対象外とした。

(6) 例文が複数の意味にとれる場合

例文が複数の意味にとれる場合の例を例 12 及び例 13 に示す。

例 12 : 職場の不満から彼は家族の者に 当たった .

対訳 12 : He was hard on his family because of complaints he had about his job .

ALT 訳 12 : He corresponded to the person of his family from the discontent of a place of work .

例 13 : 彼が転んだ彼女を 起こした .

対訳 13 : He picked her up after she fell down.

ALT 訳 13 : He woke up the woman who fell down.

例 12 は、対訳の通り『彼が家族に当たり散らした』という意味である。しかしパターンの該当部分『彼は家族の者に当たった』だけに注目すると、対訳の意味とは別に、「彼は家族の関係にある」という意味にも解釈できる。よって『N1 が N2 に当たる』は、文

によって『be hard on ~』と『corresponded to ~』の二つの訳語を使い分ける必要がある。ところがその使い分けには、この文の因果関係に注目する必要がある。一般に因果を表す文は、原因と結果をそれぞれ節で表すため複文となる。そのため単文レベルのパターンでは対処できない。これは現在の結合価文法による訳し分けの限界だと言える。

5.1.3 動詞のまとめ

以上より、動詞の訳語選択において結合価文法は有効であり、今後パターンの拡充などにより、結合価文法を最大限に活用できる環境が整えば、現状の89%の正解率から98~99%の正解率を得ることができるようになると推測できる。しかし『(8) 複数の意味にとれる文』については結合価文法だけでは訳し分けるは不可能である。

5.2 名詞の訳語選択に付いての考察

5.2.1 名詞の訳語選択への有効性

表4の結果よりALTの結果の方が、デフォルトの訳語の結果よりも6%良い値となっている。したがって結合価文法により、名詞の訳語選択に多少効果があると言えるが、動詞の場合のように、デフォルトの場合と比べて顕著な差は見られなかった。これは、IPAL辞書の名詞の約5割が多義を持たないため、デフォルトの正解率が最初から高いことと、結合価文法は動詞の訳語選択を目的として開発されたものであることが原因と考えられる。

5.2.2 訳語誤選択の原因と解決方法

動詞の訳語誤選択と同様に、ALTが正しい訳語を選択できなかった原因を、評価が×となった例文92文について調査した。結果を表6に示す。

ALTでは名詞の訳語選択において表6の問題を含む。動詞の場合と同様に、これらの問題が解決されれば、結合価文法により正しい訳語を選択できるようになると考えられる。以下にそれぞれの問題の解決方法を考察していく。なお、各原因に該当する例文は巻末の付録に記載する。

表6: 正しい名詞訳語が出せなかった原因

	訳語選択失敗の原因	割合
1	パターンが登録されていない場合	39% 36文
2	パターンの照合に失敗した場合	4% 4文
3	パターンの格要素の制約が弱い場合	17% 16文
4	対象の名詞がパターンの要素外の場合	8% 7文
5	形態素解析に失敗した場合	5% 5文
6	慣用表現	24% 22文
7	見出し語の名詞が辞書未登録の場合	2% 2文
	合計	92文

原因(1)(2)(5)(7)

原因「(1)パターンが登録されていない」「(2)パターンの照合に失敗」「(5)形態素解析に失敗」「(7)見出し語の名詞が辞書未登録」については、5.1節の動詞の考察と同じ種類の問題である。しかし動詞の場合と同様にパターンの辞書への追加、パターン照合方式の改善などにより、解決ができるとは限らない。これは以下に述べる原因(3)と(4)の問題が原因である。

(3) パターンの格要素の制約が弱い場合

日本語文に対応するパターンにおける格要素を制約する意味属性が『* (指定なし)』や『主体』などの上位の階層にある場合、訳語が一意に決まらない場合がある。具体例を例14及び例15に示す。

例14：彼は相手に意向を質した。

対訳14：He asked the other party of their intention.

ALT訳14：He asked his partner about a mind.

例15：彼の体は薬で損なわれている。

対訳15：His body has been ravaged by the drug.

ALT訳15：His body has been ruined with medicine.

例14では「N1(主体)がN2(主体)にN3(抽象)を質す」というパターンに対応している。例14の訳し分け対象の名詞「意向」には『意図』と『思想』の二つの意味属性が定義されている。しかしこの二つの意味属性は共に『抽象』配下である。よってこのパターンにおいて「意向」の意味属性は一意に決定できない。したがって、大きく訳語を絞り込むことはできても一意に決定できない。

この原因はパターンが用言の訳し分けを目的に作られたのが原因と考えられる。つまり用言の訳語が同じパターンの場合、格要素の名詞の意味属性は、他のパターンと住み分けができる程度までしか定義されていない。よってその意味属性の範囲内での格要素の名詞の訳し分けは保障されていない。この問題の解決方法として「N1(主体)がN2(主体)にN3(意図, 思想)を質す」と、意味属性の制約を強めたパターンを追加することが挙げられる。意味属性の制約を強くしたパターンの追加により、文意に合わせてパターンの意味属性を選択できるようになると考えられる。しかし入力文の情報から文意を判断し、正しい意味属性のパターンを選択することは現在の結合価文法では困難と考えられる。

(4) 対象の名詞がパターンの要素外の場合

対象の名詞がパターンの要素外の場合の例を例 16 及び例 17 に示す。

例 16：逃亡した男は地方で生き延びている。

対訳 16：The man who escaped is surviving in the countryside.

ALT 訳 16：The man who escaped has survived in a district.

例 17：私は濡れた服をストーブのそばで乾かした。

対訳 17：I dried my wet clothes next to the heater.

ALT 訳 17：I dried with the buckwheat noodle of a heater clothes that got wet.

文中で対象の名詞が、パターンの格要素以外で用いられている場合には、結合価文法による訳し分けはできない。パターンは用言と格要素の意味的關係が記述されているので、パターンに含まれない格要素の名詞については、意味属性によって訳語の絞り込みができない。例 16 では、パターン「N1(主体, 動物)が生き延びる」が適用されている。しかし対象の名詞「地方」を含む格要素「地方で」は、パターンに含まれていない。したがってこの場合、パターンに記述されている意味属性によって訳語を絞り込むことができない。

この問題もパターンが動詞の訳し分けを目的に開発されたものであるということが原因と考えられる。現在のパターンは訳し分けにおいて、最低限必要な格(必須格)と意味属性しか定義していない。よって動詞の訳し分けの観点から重要でない格の名詞にはパターンが適用できない。

この問題の解決の可能性としては、不足している格をパターンに追加する方法が挙げられる。例 16 ではパターンを「N1(主体, 動物)が N2(場所)で生き延びる」とすれば名詞『地方』も意味属性による制約を受け、訳語を絞り込むことができるようになる。しかし、同じように名詞の訳し分けのために他のパターンにも格を追加していくと、格の増加によりパターンが複雑化する。それによりパターンの誤選択が増え、動詞の訳し分けが困難になると考えられる。よって格要素を追加する際に必須格ではなく、任意格として追加する方法を考える。任意格は、パターン選択の際には考慮に入れず、パターンが決まった後、例文に任意格に対応する格があればその格に対して適用する。

例 16 の例ではパターンを「N1(主体, 動物)が [N2(場所)で] 生き延びる」([] 内は任意格とする) とし、任意格 [N2(場所)で] はパターン選択の際には考慮に入れない。そしてこのパターンが選択された後、格『地方で』に任意格 [N2(場所)で] が当たるので、名詞『地方』を意味属性 <場所> で制約する。

以上の方法で今後、パターンによって現在よりも名詞の訳語選択の精度を高めることができると思う。

(6) 慣用表現

慣用表現の例を例 18 及び例 19 に示す。

例 18：彼女は亭主を尻に敷いている。

対訳 18：She dominates her husband.

ALT 訳 18：She is spreading her husband out on the back.

例 19：彼は骨がある。

対訳 19：He has backbone.

ALT 訳 19：He has a skeleton.

慣用表現については、パターンを登録することにより、確実に対象の名詞を含むパターンができるので、訳し分けに効果が有ると考えられる。

例 18 について言えば、「尻に敷く」の部分が慣用表現である。慣用表現をパターンにする際には「N1(人)がN2(人)をN3(身体・部位)に敷く」とは記述せず、「N1(人)がN2(人)を尻に敷く N1 dominates N2」と字面で登録するため、前述した意味属性の制約の強さの問題や、パターンの格要素に含まれない、と言った問題を考慮する必要がない。よって慣用表現についてはパターンの追加によって確実に改善が見込まれると言える。

5.2.3 先行研究との比較

名詞の訳語選択については、IPAL 辞書の名詞における「意味的用法」に関する知識分類を使用した訳語選択の可能性について机上で検討した研究がある [15]。この研究では、意味属性が決定することにより、IPAL 辞書の多義を持つ名詞をどこまで訳し分けできるかを調査している。その調査によると、名詞は表 7 のように分類される。各分類の名詞の出現割合は、「訳語多義なし」56.4%、「一意に絞り込み可能」24.0%、「場合により可能」3.5%、「絞り込み可能」10.5%、「不可能」5.7%となっている。本研究の結果をこの分類に当てはめ、それぞれの正解率を調査した。その結果を表 7 に「訳し分け精度」として示す。結果を見ると、先行研究において意味属性により訳し分けが「不可能」、「絞り込み可能」までと考察された見出し語については、他に比べ誤り率が高くなっており、先行研究の考察とほぼ一致する。

表 7: 意味属性による訳し分けの可能性と実測値

分類 (参考文献 13 より)	各分類の訳し分け精度		
			×
訳語多義なし	71%	25%	4%
一意に絞り込み可能	56%	32%	12%
場合により可能	49%	42%	9%
絞り込み可能	44%	40%	15%
不可能	52%	32%	16%

5.2.4 名詞のまとめ

名詞の訳語選択において、結合価文法は顕著な成果をあげることはできなかった。その原因としては、現在使用されているパターンが用言を中心としたものであることが挙げられる。つまりパターンは用言を訳し分けるために、用言に係る格要素と最低限の意味属性しか定義していない。したがって、その定義外の名詞の訳語選択についてはパターンが適用されないため効果がない。

以上の理由から 5.2.3 節で考察した原因が改善されたとしても、全ての名詞が結合価文法によって訳し分けができると考えられない。しかし「(6) 慣用表現」は、パターンの登録によって確実に解決できる。したがって、現状の 91%の正解率から 94%の正解率を得られるようになると推測される。

なお、現在の用言中心のパターンでは上記の 94%までの精度向上しか見込めないが、今後は名詞の訳語選択の観点からパターンを改良することによる解決方法が考えられる。具体的な改良の方法は、パターンの意味属性をより詳しく定義することと、必須格でない格を名詞の訳し分けのために任意格として追加する方法が挙げられる。

6 まとめ

本稿では、単文において結合価文法の動詞と名詞の訳し分けの精度を調査した。具体的には、IPAL 辞書の見出し語に登録されている動詞と名詞についての例文を ALT-J/E で翻訳し、その結果を評価する方法で調査を行なった。ALT-J/E の使用により、パターン決定の際に頻度情報が用いられ、人手で一意に決定できないものについても正解する場合が有ったと考えられる。しかしその結果も訳語選択の根本に結合価文法を用いているので、それも含めて結合価文法の訳語選択能力の評価とした。

その結果、特に動詞についてはデフォルトの訳語と比べ 34% の精度の向上が見られた。しかし名詞はパターンの性質などから顕著な効果は見られず、6% の向上にとどまった。また失敗例の分析から、パターンの拡充などによって、結合価文法の特徴を限界まで用いることができれば、動詞においてはさらに精度向上が期待できることがわかった。しかし、名詞においては結合価パターンが動詞の訳し分けを目的に作られた経緯から、現在のものでは名詞の訳し分けが困難であることがわかった。しかし現在のパターンにさらに詳しい意味属性の定義や任意格を追加することにより、名詞の訳し分け精度を向上させられる可能性があることがわかった。以上の考察から現在の結合価パターンにおける限界は動詞で 98 ~ 99%、名詞で 94% と推測される。

今後は本研究で対象としなかった IPAL 辞書以外の動詞、名詞についても調査を行ない、結合価文法の効果をより詳しく検証する必要があると考えられる。

謝辞

本研究の作成にあたり，御指導頂きました徳久助手，池原教授，村上助教授に厚くお礼申し上げます．

また本研究で用いた「ALT-J/E」および言語関係の資源はNTTとの共同研究の関係から使用させて頂きました．ここにお礼申し上げます．

参考文献

- [1] Allen James, “Natural Language Understanding”, The Benjamin Cummings, 1995.
- [2] 麻野間直樹, 中岩浩巳, “目的言語の単語共起情報を利用した訳語選択と未知語の訳出”, 言語処理学会第5回年次大会論文集, pp.442-448, 1999 .
- [3] 野美山浩, “目的言語の知識を用いた訳語選択とその学習性”, 情報処理学会自然言語処理研究会, 91-NL-86, pp.1-8, 1991.
- [4] 鈴木克志, 太細孝, “日英機械翻訳における共起表現の扱い”, 情報処理学会自然言語処理研究会, 91-NL-82, pp.1-8, 1991.
- [5] 角田達彦, 田中英彦, “英語名詞の多義性解消における文脈としての場面情報の評価”, 自然言語処理, Vol.3, No.1, pp.3-27, 1996.
- [6] 北村博, 荻野紫穂, “日英翻訳における連体修飾句の訳し分け”, 情報処理学会自然言語処理研究会, 90-NL-75, pp.1-8, 1990.
- [7] 柏野(桑畑)和佳子, 本多啓, “IPAL 名詞辞書における多義構造の記述”, 情報処理学会論文誌, Vol.39, No.9, pp.2603-2612, 1998.
- [8] 桑畑和佳子, 橋本三奈子, 青山文啓, “IPAL 名詞辞書による多義性解消のためのコーションの分析”, 情報処理学会論文誌, Vol.39, No.6, pp.1925-1934, 1998.
- [9] 水谷静夫, 石綿敏雄, 荻野孝野, 賀来直子, 草薙裕, “文法と意味 I”, 朝倉書店, 1983.
- [10] “計算機用日本語名詞辞書 IPAL 解説編”, 情報処理振興事業協会技術センター, 1996.
[<http://www.ipa.go.jp/STC/NIHONGO/IPAL /ipal.html>]
- [11] 池原悟, 宮崎正弘, 白井諭, 横尾昭男, 中岩浩巳, 小倉健太郎, 大山芳史, 林良彦, “日本語語彙体系 5 構文体系”, 岩波書店, 1997.
- [12] 池原悟, 白井諭, 小倉健太郎, “言語表現体系の違いに着目した日英機械翻訳試験項目の構成”, 人工知能学会論文誌, Vol.9, No.5, pp.569-579, 1994.
- [13] 長谷川潔, アラン・ターニー, 橋本光郎, “PROCEED”, 福武書店, 1998.

- [14] 吉田真司, 池原悟, 村上仁一, “入力文に対する結合価パターン対の選択方法について”, 言語処理学会全国大会報告, B2-6, pp.299-302, 2002.
- [15] 桐澤洋, 池原悟, 村上仁一, “日英機械翻訳における名詞の訳語選択”, 言語処理学会, B1-6, pp.55-58, 2000.
- [16] 金出地真人, 池原悟, 村上仁一, “結合価文法による動詞の訳語選択能力の評価”, 情報処理学会第 63 回全国大会 Vol.2, pp.267-268, 2001.
- [17] 金出地真人, 徳久雅人, 村上仁一, 池原悟, “結合価文法による動詞と名詞の訳語選択能力の評価”, 自然言語処理研究会 2003-NL-153-16, pp.119-124, 2003.
- [18] 金出地真人, 徳久雅人, 村上仁一, 池原悟, “結合価文法による動詞と名詞の訳語選択能力の評価”, 言語処理学会論文誌 (投稿中) .