

日英文型パターンにおける節の変数化条件について

徳久雅人 *1 村本奈央 *2 池原悟 *1 村上仁一 *1

*1 鳥取大学工学部 知能情報工学科

*2 NTT アドバンステクノロジー

E-mail: {tokuhisa,ikehara,murakami}@ike.tottori-u.ac.jp E-mail: nao@nlp.ntt-at.co.jp

アブストラクト 等価的類推思考の原理に基づく機械翻訳方式を実現するためには、原言語および目的言語において、意味的に類似する表現を対応付けたデータベースが必要になる。そのために表現を文型パターンの形式で記述している。文型パターンは、字面、変数、関数、制御記号で構成されている。変数には、単語、句、節、という3つのレベルがある。中でも節レベルの変数を表現文より作成することは、節が言語構造上複雑であることから、不安定になりがちである。そこで本稿では、文型パターンを利用する立場から変数に対する要請をまとめ、その条件の下で文型パターンを作成することを試みる。その結果、節レベルの文型パターンを作成する手順について1つの提案をすることができた。

キーワード 文型パターン, 節変数化条件, 作成手順, 決定木, 拡張結合価パターン

1 はじめに

従来の機械翻訳では、要素合成法に基づくため方式的な限界にきている。原言語と目的言語の表現構造にのみ着目した構造変換の方式であることが原因である。

これに対し、意味的に等価な変換を考慮した翻訳方式が提案されている。その1つに多段翻訳方式があり、日英機械翻訳システム ALT-J/E に実装されている。この方式では、主観世界と客観世界の意味をそれぞれとらえ、訳文生成時にそれらを合成する。特に客観世界は、結合価文法を基礎として解析される。そのための意味体系は、[1] にまとめられている。しかしながら、この方法は単文を対象としており、重文および複文についてはカバーできていない、という問題がある。

これに対して、池原らは、等価的類推思考の原理による機械翻訳方式を提案している [2]。それは、原言語の表現の構造を、ある概念項を経由し意味的に等価といえる目的言語の表現の構造に写像を行うものである。そこでは、表現構造は文型パターンでとらえようとしている。文型パターンは、字面、変数、関数、制御記号などで構成される [3]。さらに、変数は、単語、句、節の3レベルがあり、節レベルほど汎用性の高いパターンとなることが期待されている。

この翻訳方式で日英機械翻訳を行うためには、意味的に対応のとれた文型パターンの対を収集することが重要になる。日英対訳コーパスより文と文の対応が得られるので、それらの文から単語、句、節について変

数化を行うことで、文型パターン対を得ることが出来る。しかし、節については、単語や句よりも言語構造が複雑であることから、節変数を用いた文型パターンの作成は、人手で行うとしても安定性を保つことが難しい。

そこで、本稿では、対訳から節レベルの文型パターンを作成するための作業方針を求めることを目的とする。具体的には、次の手順で議論を展開する。まず、第2章で、節変数を利用する立場から、節変数に対する要請をまとめる。第3章では、対訳に節変数を対応させ、節の部分を既存の翻訳機で訳出する。そして、その結果が節変数に対する要請を満たすか否かを調査し、不具合のある事例については対策を検討する。そして、対訳から節変数を含む文型パターン対を作成する手順を提案する。次に第4章では、節変数化の汎用性を向上させるための考察を行い、第5章でまとめを述べる。

2 節変数への要請

文型パターン対を利用する立場には次の3つがある。

- 文型パターン対を運用する立場
(例. 翻訳時に適切な文型パターン対を選択)
- 日本文を照合する立場
(例. 日本文における表層構造の解析)
- 英文を生成する立場

(例: パターン要素の訳出と文への合成)

まず、これらの立場から、節変数への要請をまとめる。

2.1 文型パターン運用の立場から

本研究で目指している翻訳方式に悪影響を及ぼさないために、文型パターンは少なくとも次の条件が守られなければならない。

(原則1) 日本文の特徴をとらえた日本語文型パターンが選択できること

(原則2) 対応する英語文型パターンに基づき英文が生成できること。特に、変数間は基本的には独立していること。

この考え方から、文型パターンの変数には次の条件がたてられる。

[条件1: 文型特徴の明示] 文型選択に関わる表層的特徴が変数化により隠れないこと

[条件2: 局所翻訳の保証] 変数化された内容は、それだけで正しく訳出できること

[条件3: 変数内容の独立] 他から変数化された内容が参照されないこと

[条件1]に関して「～しはじめる」などの様相表現、「全く～ない」などの係り結び、および「非常に～なので～できない」などの副詞や接続詞を含む表現、については、注意が必要である。これらの表層的特徴を効率良く扱うために、構文を指定する関数に置き換えた表現が検討されている(例「～できない $V1.dekiru.nai$ 」)[3]。

[条件3]の判断は、対訳の一部を変数化することにより、その変数化対象外の部分に生じる影響についても注意を払う必要がある。

2.2 日本文照合の立場から

日本文と日本語文型パターンを照合する際、日本文の字面、および、形態素情報が照合対象となる。したがって、これらの表層情報から節変数を定義する。

[条件4: 表層構成要素] 以下の格要素および用言から構成されること

- 名詞句 + 「は, が」格が少なくとも1つ存在
- 名詞句 + 「を, から, より, へ, まで, に, で」格が0個以上存在
- 動詞・形容詞・形容動詞がただ1つ存在

2.3 英文生成の立場から

入力の日本語文と文型パターンを照合すると、変数には、入力文の対応する部分がバインドされる。バインドされた部分のみを翻訳することを局所翻訳という。

節変数の局所翻訳は、単文相当であることから、結合価パターンを用いた翻訳が有効である。そこで、節変数の値は、文中の対応個所が標準表記、すなわち、各単語は見出し語に、そして、述語となる活用語は終止形に変形されて割り当てられる。たとえば「太郎はやまに登った」という文に対して、文型パターンを照合した結果を次に示す。

(単語レベル文型パターン) $N1$ は $N2$ に $V3$ た:

(結果) $N1 =$ 太郎, $N2 =$ 山, $V3 =$ 登る

(句レベル文型パターン) $NP1$ は $VP2$ た:

(結果) $NP1 =$ 太郎, $VP2 =$ 山に登る

(節レベル文型パターン) $CL1$ た:

(結果) $CL1 =$ 太郎は山に登る

次に、局所翻訳を行う。上記の例に対応する英語文型パターンに「 $CL1.past$ 」があるとすると、次の順に局所翻訳が行われる。

(1) $CL1 =$ 「太郎は山に登る」の翻訳により、 $CL1 =$ 「Taro climbs a mountain.」を得る

(2) $past$ 関数により「Taro climbed a mountain.」を得る

基本的に以上の過程を想定すると、英文生成の立場からの要請は次の通りとなる。

[条件5: 標準表記のバインド] 変数の値は、標準表記(単語は見出し語、述部の活用語は終止形)をバインドする

[条件6: 変形操作の許可] 変数の訳出に変更を加えるための指定を変数に付属させても構わない

3 節変数化作業の試行と分析

第2章では、文型パターンを利用する立場から条件を示した。本章では、文型パターンを作成する立場から、作成作業について検討する。

3.1 分析の目的と方法

本分析では、文型パターンを作成する際の注意点を明確にすることを目的とする。

具体的には、まず〔条件1,3,4,5,6〕に着目して、対訳から節レベル文型パターン対を作成する。

次に〔条件2〕を確認するために、ALT-J/Eにより局所翻訳を行い、その翻訳結果(ALT訳と呼ぶ)と対訳(理想訳と呼ぶ)を比較する。

最後に、条件が不成立となった原因について考察し、文型パターン作成に関する注意点を求める。

3.2 節事例の収集

意味類型用日英例文ファイル[4]より抽出した「～と～」という構文の対訳102件を対象としたところ、79個の節が抽出できた。

収集事例の例を以下に示す。

日文：彼はそのニュースを聞くとすぐさま帰った。

英文：He left the instant he heard the news.

CL1：彼はそのニュースを聞く

CL2：×

CL1のALT訳：He hears that newscast.

CL1の理想訳：He hears the news.

「と」の前後の部分がCL1およびCL2の候補である。CL1は節事例であり、CL2は「はが格」が無いため節事例とならない。

なお、日本語では主格の省略が多発することから、厳密には〔条件3〕が成り立たないことが多い。たとえば、上記事例の場合、CL1と節変数化することにより「彼」という情報が抽出できなくなるため「帰った(left)」の主語が不明になる。そこで、主格に関しては〔条件3〕の解釈をゆるめて節事例を検討する。

3.3 局所翻訳の妥当性判定

3.3.1 局所翻訳の妥当性判定基準

妥当性の判定は4段階で行う。

：理想訳とALT訳が完全に一致する場合

：理想訳とALT訳が同義とみなせる場合

：理想訳とALT訳が不一致であるが、知識補強により訳出が見込める場合

×：理想訳とALT訳が不一致であり、訳出が簡単には見込めない場合

「×」以外の判定は、局所翻訳に成功しているとみなす。

3.3.2 妥当性判定結果

表1に、4段階評価の結果を示す。約73%は〔条件2〕を満たしていることがわかる。

表1: 節事例の訳出評価結果(格要素は補完せず)

判定				×
件数(%)	21 (27%)	21 (27%)	16 (20%)	21 (27%)

各判定ごとの内訳を表2から表5にまとめる。なお、これらの内訳には、第4章で追加する節事例についての内訳も含まれているが、傾向を分析する上で全く問題は無い。

3.4 ×の節事例に対する考察

×にならなかった節事例は〔条件1~6〕に準じることで、節レベル文型パターンが作成できたといえる。

一方、×になった節事例は、それらの条件では不都合が生じたと考えられる。

そこで、×になった事例を考察し、文型パターン作成の手順に関して考察を行う。

3.4.1 構文選択の誤り

形式主語のit 形式主語のitを含む文が理想訳の場合、itが少なからず他の要素と関わりを持つため節変数化ができない(〔条件3〕が不成立)。対訳を以下に示す。

表 2: 判定「 」の内訳

分類	一致とみなした例
冠詞の差異	「 a/an 」, 「 the 」
数の差異	単数形と複数形
敬称の有無	「 Mr. 」 など
「 人 」の訳語	「 one 」, 「 people 」 など
「 その 」の訳語	「 the 」, 「 that 」

表 3: 判定「 」の内訳

分類	同義とみなした理想訳と ALT 訳
同義語句	「 目が覚める 」 = “wake”, “awake” 「 着く 」 = “get to”, “arrive at” など
平易化表現	「 彼は私の姿を見ると ~ 」 (理想訳) “he saw me” (ALT 訳) “he saw my figure”

日文：その光景のことを思うと 気分が悪くなる。

英文：It makes me sick when I think of the sight.

下線部の局所翻訳結果は「 We are unwell 」である。正しく訳出するためには、句・節レベルの格要素(相当)を許容するという拡張を、日本語語彙大系の結合価パターン対に施して「句節レベル結合価パターン対」を作成すればよいだろう。すなわち、主節の用言は字面で残し、その他の格要素(相当)を変数化する。

(句節レベル結合価パターン対)

日バ：N1はVP2と気分が悪くなる

英バ：it makes N1 sick when N1 VP2

変数制約：N1(4人), VP2(32思考動作)

なお、このパターン対には、日本語語彙大系を作成した際と同様に、必須の格要素を対訳に補って変数化している。

以上から「節変数化の候補が主節であり、独立性が無い場合、句節レベル結合価パターン対の作成を試みる」という作業手順が考えられる。

動詞文、形容詞文、there 構文などの選択 次の対訳では、述語が動詞である。一方、下線部の ALT 訳は「 He is¹ ashamed 」という形容詞述語である。

日文：それを考えると 彼は恥ずかしくなった。

表 4: 判定「 」の内訳

分類	正しい訳出のために必要な強化内容
用言	結合価パターンの登録 不定詞型訳出規則の登録 自/他動詞の使い分け規則の登録 結合価パターン選択アルゴリズムの改善
体言	単語の登録 名詞句の訳出規則の登録 数の訳出規則の改善
その他	助動詞・前置詞・擬音語・擬態語・はが構文・様相つき「だ文」・特定述語などの訳出

表 5: 判定「 × 」の内訳

分類	内容
構文選択の誤り	節外の句・節を受ける it の訳出 there 構文と動詞文の選択 動詞文と形容詞文の選択
訳語選択の誤り	様相表現を要する訳出 時間表現語の訳出 日本文の節内非対応単語の訳出 短絡表現の訳出(意味補完を要する訳出)
その他の誤り	「も格」の対応判別 代名詞の訳出 意識

英文：He blushed at the thought of it.

正しく訳出するために、英文生成部に対する生成条件を文型パターンから与えなければならない。そのために「構文指定」を行う。以下にその具体例を示す。

(構文指定付き文型パターン対)

日バ：NP1を考えるとCL2くなった

英バ：CL2.verb.type.past at the thought of NP1

これにより、CL2の訳出は、動詞型による訳出に限定できる。「there 構文」についても同様に対処できるようになるだろう。

以上より「訳出の構文を具体的に指定すること」が必要である。

3.4.2 訳語選択の誤り

様相表現を要する訳語選択 次の対訳では、下線部の訳出に様相表現が必要である。

日文：カメラ一体型ビデオが、5分以上使われないでいるとオートシャットオフ機能が電源を切ります。

¹ 「なくなった」を含めると「became」となる

英文：An automatic shutoff function powers down the camcorder after it sits idle for more than five minutes.

下線部は、否定の様相表現を外すと「(カメラ一体型)ビデオが5分以上使われる」となる。このALT訳は「A video is used over five minutes」となる²。ALT訳を基準に英語文型パターンを作成すると次の文型パターン対が得られる。

日パ*：CL1 ないでいると CL2

英パ*：CL2 after CL1.not

しかしながら、理想訳と比べると、CL1の肯定/否定の表現に差が生じる。この訳出の差を埋めるためには、様相表現を節変数の中に入れておく必要がある。ここで、変数のバインド値は「標準表記」とするという〔条件5〕と衝突する。

さて、日本語語彙大系における結合価パターンには、様相表現を必要とするものがある。たとえば「N1がN2を遊ばせる」である。このような用言を節変数と照合させるには〔条件5〕に対して次の例外を導入する必要がある。

〔例外1：標準表記のバインドに対する例外〕結合価パターンに付随する表層表現は、標準表記の一部としてバインドする。

さらに、「様相表現との共起が重要」とされている用言は、日本語語彙大系に収録されていないことがある。そこで、表現を拡張した「拡張様相型結合価パターン対」を対訳に応じて追加する。そして、局所翻訳において、それが使われるという前提の下で、文型パターン対を作成する。

(拡張様相型結合価パターン対)

日パ：N1が使われないでいる

英パ：N1 sit idle

変数制約：N1(962 機械)

(文型パターン対)

²受動態を能動態に戻してALT-J/Eで翻訳しても、主格が無いため受動態の英訳が出力される。

日パ：CL1とCL2

英パ：CL2 after CL1

以上より、「様相表現のあることで英文の本動詞が定まる場合には、拡張様相型結合価パターン対を定義し、それが節に対応するという考えの下で、文型パターン対を作成する」という作業手順が考えられる。

時間表現の訳出方法 次の対訳では、時間表現の訳出に工夫が必要である。

日文：彼女の態度から考えると彼女はそのことを前から承知していたようだ。

英文：Judging from her attitude, she seems to have been aware of it for a long time.

下線部は、述語の用言および必須の格要素に注目すると、「彼女はそのことを承知する」という部分と「前から～ていたようだ」という部分とに分かれる。前者のALT訳は「She consents to that thing」である。このALT訳を英文に近づけるためには、後者に注目して、現在完了形および継続期間を組合せて「seem to have ~ for a long time」という構文に変形する必要がある。これは多段翻訳方式の考え方である。

ここで、文型パターンによる翻訳では、多段翻訳方式の考え方をを用いる段階、すなわち、主観的表現を抽出する段階を明確にしておく必要がある。

仮りに文型パターンを照合する際の重要な表層的特徴として主観的表現を位置付けるならば、次のように表層的特徴を明示した文型パターン対を作る。

(包含要素指定付き文型パターン対)

日パ：NP1から考えるとCL2.rm(前から)ていたようだ

英パ：Judging from NP1, CL2.seem.to.prfct³ for a long time

この文型パターン対では、関数 $rm(arg)$ が用いられている。この関数は、変数CL2のバインド値より arg と一致する部分を取り除く働きがあり、かつ、取り除くことができた場合のみ、照合が成立するものである。こ

³prfctは現在完了形の意

のような文型パターン対を「包含要素指定付き文型パターン対」と呼ぶことにする。

逆に、文型パターンを照合する際の重要な特徴としない、すなわち「～から考えると～」という構文と「前から～ていたようだ」という構文は無関係とみなすならば、次のように文型パターン対を作る。

(文型パターン対)

日パ：NP1 から考えると CL2

英パ：Judging from NP1, CL2

しかし、この場合〔例外1〕に該当せず、節変数にはモダリティがバインドされない。そのため、訳出の最終結果には、モダリティが反映されない。

そこで、多段翻訳方式を利用する。図1を参照しながら説明する。文型パターンの節変数CL2と主観抽出規則は、原文において対応する個所が重なっている。このことから、CL2に主観表現が関わっていることが判断できる。そこで、CL2の局所翻訳において多段翻訳方式により、的確な訳出が可能になる。

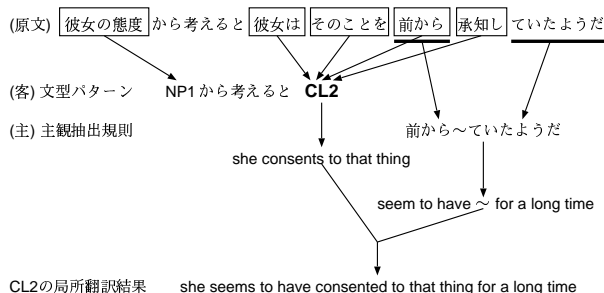


図1: 多段翻訳方式による節の局所翻訳

以上より、「対訳に時間表現が含まれている場合、文型パターンに関わることが本質であるならば、文型パターンに包含要素指定つき文型パターン対を作成し、そうでなければ、局所翻訳における多段翻訳方式が働くことを期待して文型パターン対を作成する」という作業手順が考えられる。

ただし、対訳から時間表現が構文に本質的に影響するか否かを判別することは問題として残される。

節内非対応単語の訳出 英文の単語が日本文の単語に対応しない場合がある。幾つか対訳の例を示す。

日文：田舎は都会に比べると 文化が遅れている。

英文： Cultural things are somewhat behind in the country compared to in the big city.

— 「somewhat」に相当する日本語単語が存在しない。

日文：私が出かけようとしていると 彼がやって来た。

英文： I was about to go out, when he came to see me.

— 「to see me」に相当する日本語単語が存在しない。

日文：彼女が来ると ほかの者はみな霞んでしまった。

英文： All the people present were outshone by her appearance.

ALT 訳： All the people of other are eclipsed.

— 「present」と「ほかの(者)」が直接的には対応しない。

前者は、ALT 訳では「A culture is underdeveloped」となる。これは結合価パターン対「N1(文化,1002,1155,458)が遅れる N1 be underdeveloped」により訳出される。対訳の通りの訳出には、「Cultural things be somewhat behind」を直接定義する他にない。

中者を訳出するためには、日本語の従属節から「私の所」に注目されること、主節において「やって来る先」が「私の所」と判定されること、そして、結合価パターン対「N1がN2の所にやって来る N1 come to see N2」があること、が必要となる。また、節間で情報を授受しているため〔条件3〕が成立しない。したがって、節変数化はできない。

後者は、集合を表す関係について〔条件3〕が成立しないため、節変数化はできない。つまり、日本文では「彼女」が文頭に書かれた後に「ほかの者」となり、英文では「all the people present」が書かれた後で「her」となる。集合のうち、1つあるいは幾つかを指定した後ならば「other」の意味が理解できるが、それらの指定なしに「other」は意味をなさないだろう。

3.5 節変数化の基準

前節での考察より、結合価パターン対および文型パターン対の記述に拡張が生じた。次項ではまずそれらをまとめる。次に、節レベル文型パターンの作成手順をまとめる。

3.5.1 結合価パターンと文型パターンの記述能力

まず、結合価パターンの記述能力を2つの観点からまとめる(表6)。

表 6: 結合価パターンの記述能力

様相の有無	様相無し	様相表現とは無関係に、格要素と用言の共起関係だけで結合価パターンが定まる
	基本様相型	格要素、用言、および、様相表現の共起関係で定まる。[1]に収録済
	拡張様相型	基本様相型における様相表現よりも複雑なもの
格要素のレベル	名詞レベル	名詞および名詞句を含む
	句レベル	上記に加え、動詞句を含む句は格要素相当とみなす
	節レベル	上記に加え、節を含む節は格要素相当とみなす

日本語語彙大系には、様相無し・基本様相型、かつ、名詞レベルで構成された結合価パターンが収録されている。句・節レベルは、重文・複文であり、句・節レベルの文型パターンと重なる部分がある。

次に文型パターンの記述能力を変数操作に関してまとめる(表7)。

表 7: 文型パターンの記述能力(変数操作の観点から)

無操作	変数に指定が一切無関係
活用指定	用言型変数への活用形の指定
構文指定	日本語の場合「～ている」など様相の指定、 英文の場合「there 構文」、「verb_type」 など構文構造の指定
包含要素指定	変数内の特定要素の有無を指定

以前、字面関数(字面検索除去関数)という包含要素指定に類似した関数があった。包含要素指定は、その拡張型であり、字面以外(たとえば、変数)の検索と除去が可能となる。これにより、日本語のうちそのままでは変数で隠されてしまう部分を、文型パターンに明示させることができる。こうして[条件1]の実現を狙うことができる。

3.5.2 節変数化の手順

まず、基本的な節レベル文型パターンの作成基準を示す。

表層レベル

- (1) 日本語の表層構造が定義と一致すること[条件4,5][例外1]

- (2) 対応する英文も節であること

意味レベル

- (3) 節から訳出できること[条件2]
- (4) 他の部分の訳出に節の内容が使われないこと[条件3]

文型パターン運用

- (5) 節内に日英文型パターンを選択する際のキーワードを含まないこと[条件1,6]

前節で考察したとおり、より高度な節レベル文型パターン対を作成するためには、図2に示す決定木に従うことで、文型パターン対を作ることができるだろう。

startより文型パターン対を作成しようとする。まず、対訳より節変数化の対象となる部分を抽出する。日英ともに節の構造が存在するならば、文型パターン対の選択に必要なキーワードを探す。このキーワードは、最終作業の際、活用や構文の指定に使われる。キーワードのチェックを終えると、日本語を節変数CLにバインドする。このとき、様相表現を取り込むべきであるか検討する。

次に、CLの独立性を調べる。文内の節外部分と関連がある場合には、節変数化を諦める。ただし、句節レベルの結合価パターンとして用いられる可能性はある。

次に、CLを局所翻訳し訳出の失敗を調べる。失敗した原因に応じて対応がとられる。詳細は前節で述べたとおりである。

最後に、文型パターンの特徴が表れるように活用や構文などの指定を集計し、節レベル文型パターン対を登録する。

4 汎用性向上に向けての考察

[条件4]に「は格」あるいは「が格」を必須の格要素と定義している。この定義では、主格の省略の多い日本語では、汎用性の向上が期待できない。そこで、「は格」の補完を許して照合されることを想定し、汎用性の向上をねらう。

前章での102件の対訳において「は格」の補完を行い第2章の定義により節変数化を行ったところ、57個の節が得られた。第3.3.1項の結果とあわせて136個の

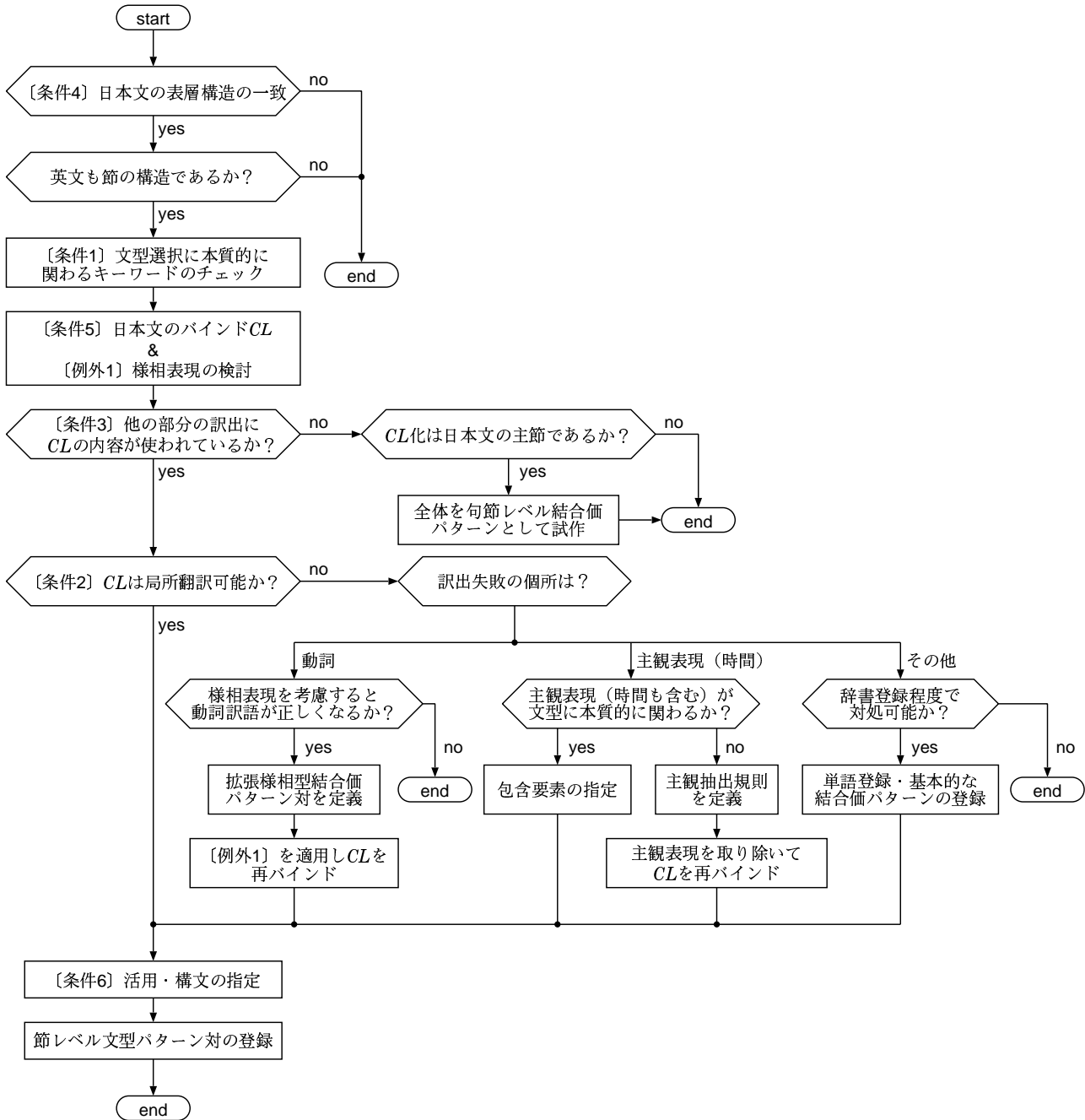


図 2: 節レベル文型パターン対作成のための決定木

表 8: 節事例の訳出評価結果(「は格」補完の場合)

判定				x
件数 (%)	12 (21%)	23 (40%)	15 (26%)	7 (12%)

表 9: 補完した「は格」の内訳(値は節件数)

	一致	不一致	計
文内要素	15	3	18
文外要素	27	12	39
計	42	15	57

節が得られたことになる。これにより節レベルの変数化の数が 1.7 倍になる。

「は格」を補完した結果に対して、ALT 訳の精度(表 8)、および、訳出誤りの原因は、第 3 章で議論したものと同様の傾向であった。

ここで問題となる点は「は格」補完の実現可能性である。そこで補完内容の内訳を調べると表 9 の通りとなった。表 9 において「文外要素」とは、補完内容が文外の要素であることを表す。つまり、翻訳時に文外情報として与えられる内容である。「文内要素」とは、文内のもう一方の節における要素で、補完した節の数を表す。一方「一致/不一致」とは、他方の節の主格と同一であるかにより分類している。

従って、粗い検討となるが、与えられた文のうち、片方の節に主格が無い場合、もう一方の節の主格と一致させることで約 74%(57 個中 42 個)の精度で補完ができる。また両方の節に主格が無い場合は、文外情報として主格のデフォルト値を利用することになるが、約 68%(57 個中 39 個)の精度で補完ができるといえる。

補完を完全に行うことは不可能であるため、主格の補完による汎用性の向上は、訳出精度の低下とトレードオフの関係になる。

5 おわりに

本稿では、等価的類推思考の原理に基づく機械翻訳方式を実現するために、対訳から節レベルの文型パターン対を作成するための基準について検討した。具体的には、以下のことを行った。

- 文型パターンを利用する立場から節変数への要請として、6 つの条件を示した。

- 「～と～」の日本語構造のある対訳 102 件について、節レベルの文型パターン対の作成を試みた。
- 作成は上記の 6 条件に基いた。その結果約 73% が適切な節変数化であった。
- 不適切であった約 27% の節事例に対して考察を行い、不適切さを排除するための作業手順を提案した。

また、日本語における主格の省略に対し、主格の補完による節変数化を試みたところ、節事例が 1.7 倍になった。ゆえに節レベル文型パターンの汎用性向上に効果が期待できる。しかしながら、簡単な補完手法では、7 割前後の補完正解率と予想されるため、訳出精度とはトレードオフの関係になる。

今後の課題は、まずは、受動態などの構文変形について検討が不足しているため、検討対象を拡大することである。次に、本作業手順に従うことで節レベルの文型パターン対が正しく作成できるのか確認することである。

また、この方法で作成した文型パターン対を用いて翻訳をするためには、主観抽出規則を用いた多段翻訳方式や、拡張様相型の結合値パターンの利用など、複雑な要素がある。これらによる英訳生成過程が妥当であるかどうか、検討することも課題として残されている。

参考文献

- [1] 池原, 宮崎, 白井, 横尾, 中岩, 小倉, 大山, 林: “日本語語彙大系”, 岩波書店 (1997).
- [2] 池原, 佐良木, 宮崎, 池田, 新田, 白井, 柴田: “等価的類推思考の原理による機械翻訳方式”, 信学技報, TL2002-34, pp. 7-12 (2002).
- [3] 池原, 宮崎, 佐良木, 池田, 白井, 村上, 徳久: “機械翻訳のための日英文型パターン記述言語”, 信学技報, TL2002-48, NLC2002-90, pp. 1-6 (2003).
- [4] “日英対訳コーパス編成ファイル” (2002).