

日本語 ベトナム語機械翻訳システム jaw/Vietnamese における翻訳実験 - 連体修飾(「の/こと」など)、86 例文に対する実験 -

田中友樹 Nguyen My Chau 池田尚志

岐阜大学工学部

1. はじめに

我々は日本語 ベトナム語機械翻訳システム jaw/Vietnamese の構築実験を行っている。機械翻訳において、名詞を修飾する連体修飾構造は言語によって違いが大きく、翻訳処理の方法は重要である。

[2]では修飾される名詞が一般名詞の場合について述べたが、本報告では修飾される名詞が形式名詞(「の/こと」)の場合についての実装と実験について報告する。

また、日英対訳集[3]から抽出した例文についての翻訳実験とそこで現れた問題の処理についても述べる。

1. 機械翻訳システム jaw の概要

jaw は、日本語から多言語への翻訳を行うパターン変換型の翻訳エンジンである。

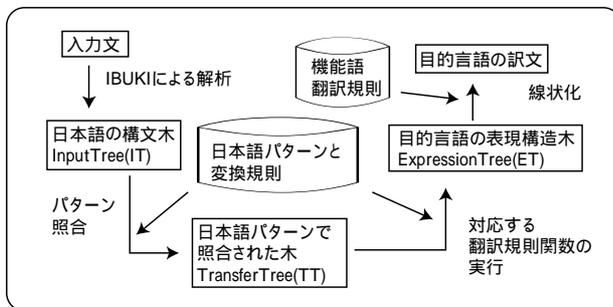


図 1 jaw 翻訳処理の流れ

2. 日本語とベトナム語の連体修飾構造の翻訳処理について

2.1 日本語の連体埋め込み修飾表現のベトナム語への機械翻訳

日本語の埋め込み文による連体修飾構造はすべて次のような構造をとる。

{ 述語の連体形 } + N 被修飾名詞
昨日あなたが会った人は私の弟です。
{ 昨日あなたが会った } + 人

日本語の連体埋め込み修飾表現は、ベトナム語では被修飾名詞が一般名詞の場合には、関係節構造と単純修飾構造をとり、形式名詞の場合にはさらに名詞節構造、副詞節構造を加えた4つの構造のいずれかで表現される。

関係節構造

- ・太郎になぐられた子供が入院しています
Đứa bé mà đã bị Taro đánh thì đang nằm viện.
子供 関係詞 過去 受身 太郎 殴る だ 進行 入院する
- ・逮捕されたのは山田社長だ
Người mà đã bị bắt là giám đốc Yamada.
の(人) 関係詞 過去 受身 逮捕する だ 社長 山田

単純修飾構造

- ・お寺がたくさんある町が好きです
(Tôi) thích phố có nhiều chùa
(私) 好き 町 ある たくさん お寺
- ・古いのはおいしくない
Cái cũ thì không ngon.
の 古い である 否定 おいしい

名詞節構造

- ・彼は仕事に夢中で帰るのを忘れた
Mê mải với công việc nên anh ấy đã quên về.
夢中 に 仕事 で 彼 過去 忘れる 帰る

副詞節構造

- ・カゼをひいたのは天気が急に変わったからだ
Sở dĩ (tôi) đã bị cảm là vì thời tiết đã thay đổi bất ngờ.
所以 (私)過去 カゼ から 天気 過去 変換する 急に

2.2 jaw/Vietnamese の埋め込み構造の処理

XN という係り受け関係があるパターンに対する照合処理は、連体埋め込み構造でなくとも、一般に (a) N の BaseTypeRule の照合

- (b) X の AdditionCWTypeRule の照合
 - (c) X の AdditionFWTypeRule の照合
- という順序でなされる。

連体埋め込み構造の場合は、X が述語で N が名詞文節である。

さらに、(d) X の BaseTypeRule の照合を行う。連体埋め込み構造は一般に修飾部分と被修飾名詞の間に格関係がある「内の関係」と、格関係がない「外の関係」に分類できる。言語によっては内の関係であるか外の関係であるかによって表現の仕方が異なる。現在のところ jaw では上記の照合処理のうち (a)(b)(c) で照合された場合を外の関係、(d) で照合された場合を内の関係と判断している。

jaw では、さらに(a)(b)(c)のパターン記述に基づかない外の関係の照合として、

- (e) <用言> <体言> の照合を設けている。
- これは V N を照合するもので、default の外の関係の照合として機能する。

以上の照合処理は目的言語に依存しない翻訳エンジンとしての処理である。

- 目的言語に依存するのは線状化関数の部分である。ベトナム語の場合の連体埋め込み修飾されている名詞の線状化処理は、被修飾名詞が一般名詞の場合は、
- (1) 形容詞 X が N を修飾する
 - (2) 名詞 N を形容詞 X が後ろから修飾する
 - (3) 名詞 N の後に関係詞をとらずに動詞 X が N を修飾する単純修飾構造となる

- (4) 名詞 N の後に関係詞 mà をとって動詞 X が N を修飾する関係節構造となる
- の4つのいずれかになる。

ベトナム語では、N が実質名詞の場合、X が形容詞であれば、X N となるか N X となるかの訳し分けが必要であり、X が動詞であれば、単純修飾構造(N X) か mà を関係詞とする関係節構造(N mà X)かのいずれかとなり、その訳し分けが必要である。

- N が形式名詞の場合は、
- (5) 先行詞 A、関係詞 mà をとる関係節構造となる
 - (6) 先行詞 A で関係詞をとらず単純修飾構造となる
 - (7) X のみの名詞節構造となる
 - (8) 接続詞 c1, c2 をとる副詞節構造となる
- のいずれかになる。これらの構造に関する選択規則と先行詞の決定規則が必要になる。

2.2.1 jaw における一般名詞の場合の処理

被修飾名詞が一般名詞の場合におけるベトナム語への処理のアルゴリズムを図2に示す。[2]

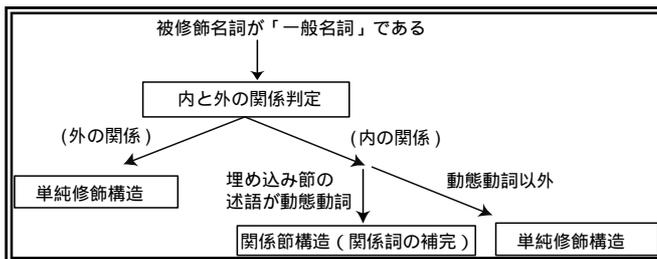


図2 一般名詞の訳し分けアルゴリズム

2.2.2 jaw における形式名詞の場合の処理

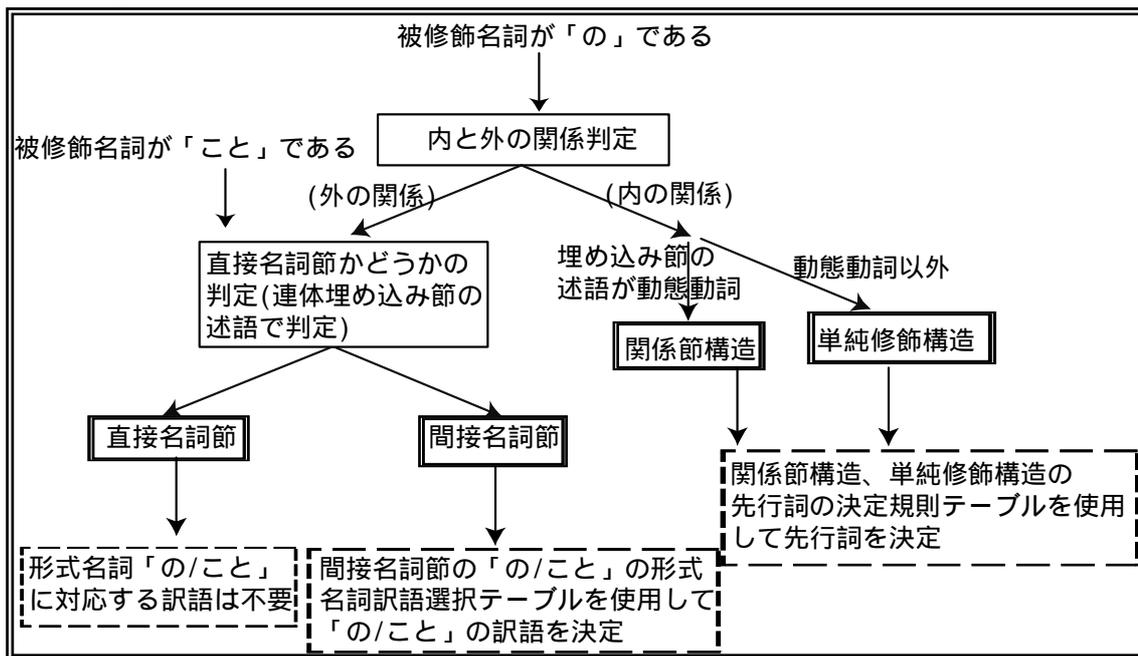


図3 形式名詞の訳し分けのアルゴリズム

被修飾名詞が形式名詞の場合の連体修飾構造におけるベトナム語への処理のアルゴリズムを図3に示す。被修飾名詞と埋め込み節述語が内の関係か外の関係のどちらであるかや、埋め込み節の動詞が

表1 形式名詞「の」の訳語を機械翻訳で決定するための判定条件の一部

順	主節における名詞Nの意味属性	主節の動詞	先行詞N
1	人 / 人名 / 不定称「だれ」	だ	người
2	物品 / 植物	だ	vật
3	無生物 / 名詞「何」 / 名詞「これ, それ, あれ」	だ	cái
4	動物	だ	con
5	国名	だ	nước
6	場所	だ	nơi
7	埋め込み節の動詞が伝達動詞の意味属性	その他	điều
8	埋め込み節の動詞の意味属性が伝達動詞以外	その他	những gì
9	静態動詞	その他	cái

動態動詞か静態動詞であるかなどの条件をみていくことにより、図3のアルゴリズムに従って、各構造に判定していく。また、表1は形式名詞「の」の訳語を機械翻訳で決定するた

めの判定条件の一部である。

2.3 連体修飾表現(形式名詞の場合)の翻訳実験

日本語の連体埋め込み修飾表現で被修飾名詞が形式名詞の場合の例文をベトナム語での表現に翻訳する実験を行った。

- (1) 花瓶を壊したのはトムです。
người mà đã làm vỡ bình hoa là Tom
 - (2) 私が住んでいるのは日本だ。
nơi tôi đang sống là nước Nhật
 - (3) 母は息子が大学に合格するのを喜んだ。
mẹ <đã> vui mừng việc con trai thi đậu đại học
 - (4) 戦争が悲惨なのを悲しんだ。
<đã> đau khổ sự chiến tranh tàn khốc
 - (5) 家が揺れるのを感じた。
<đã> cảm thấy nhà lắc lư
- 関係詞出力判定のための、内外の関係の判定が難しくその精度向上が課題となる。

3. 86 例文に対する翻訳実験

3.1 jaw/Vietnamese での 86 例文の翻訳実験

英作例文集[3]の例文 1000 文から任意に選んだ 86 例文に対して jaw/Vietnamese で小規模な実験を行った(表2)。

表2 86 例文に対する翻訳結果

実験文	正訳	誤訳	正訳率
86	70	16	81%

3.2 実験で現れた問題についての対処法

3.2.1 離合詞の処理

彼は私に全ての有り金を貸す。

anh ấy cho mượn tất cả tiền có trong tay tôi
彼 くれる 借りる 全て 有り金 私

この日本語例文で、ベトナム語訳はS+V+O+O になっている。

しかし、「全ての有り金」+「私に」という順番になると、日本語では「彼」が所有しているはずが、「私」が全ての有り金を所有していることになり、誤訳になってしまう。

この問題を解決するにあたって、「貸す」というベトナム語が、「くれる」+「借りる」で「貸す」という意味となり、正しい語順にするには、「くれる」「借りる」の間に「私」という貸す対象を置かなければならない。この理由は「貸す」は離合詞であるからである。現在のところ離合詞は「貸す」のみが該当している。

離合詞を通常の動詞と同様に扱うとするとその翻訳規則ではm_centerW(訳語の中心語を入れる変数)に離合詞の訳語を記述することになる。この場合には線状化の際にm_centerWにセットされた文字列を2つに分割し、その間に要素を挿入するという特殊な作業が必要となる。

このような特殊な作業を避けるため離合詞の翻訳規則では変数m_centerWに訳語を記述するのではなく、変数m_centerWに「離合詞を構成する文字のうち本来は動詞であった部分」の訳語を、m_accusative(目的語となるオブジェクトを指定する変数)に「離合詞を構成するも文字のうち本来は目的語であった部分」の訳語を記述することにした。このような記述方法であれば離合詞はほぼ通常の動詞と同じように扱うことが可能だからである。

次が変更した後の例文の正解の翻訳結果である。

