

統計情報と構文知識を利用した「把」文型の翻訳手法

尹大鵬 邵敏 任福継 黒岩眞吾
Dapeng Yin Min Shao Fuji Ren Shingo Kuroiwa
(徳島大学)

中日機械翻訳において、中国語の「把」文型の処理という非常に難しい問題がある。現在、いくつかの機械翻訳システムが実用化されている。しかし、現在の翻訳システムはまだこの問題をうまく解決できていない。

本研究では、中日機械翻訳に際し、「把」文型になりうる文字列を抽出して、「把」文型の自動翻訳手法を提案する。原言語の解析から目的言語の合成に至るまで、何段階かの処理が必要とされる。まず、我々は大量の教科書、科学技術文献から「把」文型の例文を収集して構文解析し、「把」文型の翻訳ルールをまとめ、最後にこのような実例から抽出したルールを中日機械翻訳実験システムに組み込む。

キーワード：「把」文型、構文解析

The translating tactics of “BA” sentence which using statistical information and Syntax knowledge

Dapeng Yin, Min Shao, Fuji Ren, Shingo Kuroiwa

There is a common but difficult problem in a Chinese-Japanese machine translation system, which is to process “BA” sentence. Nowadays, there have so many practical translation systems already. However, they can not solve the problem very well.

In this paper, during the process of the translation, we have proposed a way to handle this problem that it takes the word out automaticly from the “BA” sentence, and then divide the primitive sentence into several parts to deal with. First of all, a large number of grammar items form Chinese grammar books are collected, some elementary judgment rules are set by classifying and inducing the collected grammar items. Then, these judgment rules are put into use in actual Chinese language and are modified by checking their results instantly. Rules are checked and modified by using the statistical information from a actual corpus.

Keywords: “BA” sentence, Parsing

1. はじめに

機械翻訳に関する研究における主な問題は、構文解析と意味解析であると考えられる。特に中国語の構文解析を行う際、英語の解析に使われる方法理論をそのまま流用すると大きな問題が生じる。中国語ではいくつかの漢字が結合して複合語となる場合が多い。特に「把」が含まれる文型が多い。「把」文型は変化が多く、意味も複雑であるため、構文解析における誤り、中日機械翻訳における曖昧性と深く関わっている。中日機械翻訳においては、「把」文型を正しく処理するのは極めて重要である。

中国語の「把」文型は日本語の「他動詞文」とよく似ている。「把」文型の構文とは、本来動詞の後ろに置かれる目的語を、前置詞“把”の目的語として動詞の前に置き、この前置詞フレーズが動詞に対して連用修飾語の働きをする構造である。「把」文型構文の表す意味は事物をどう処置するかという意味を表す。

例1. 他把練習作完了。
(彼は練習をおわりました。)

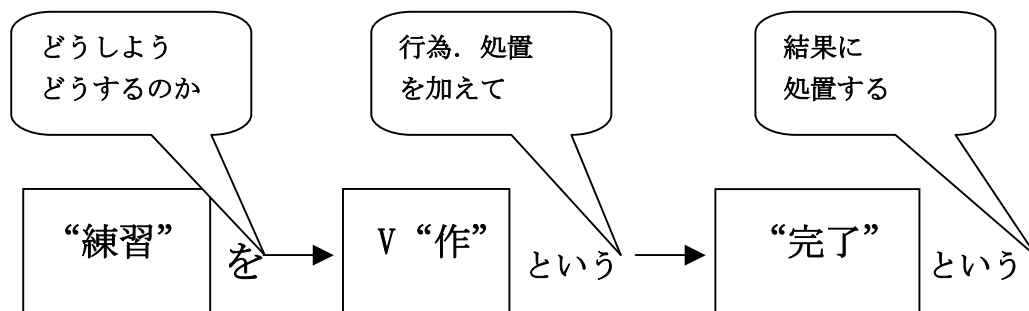


図1 「把」構文の表す意味

本稿では、まず「把」文型自動翻訳手法の流れについて述べ、その後詳細について述べる。

2. 「把」文型自動翻訳の流れ

「把」文型自動翻訳の流れを以下に示す。

1. 「把」文型の形態素解析
2. 五段モデルに変換（構文解析に対応）
3. 中国語コード列を日本語のコード列に変換
4. 語順の変換
5. 格助詞の確認
6. 時制を確認し日本語述語の時制を決定
7. 翻訳結果出力

1では入力された「把」文型の中国語文に対し形態素解析を行う。2では1の結果に基づき図2に示す五段モデルに変換する。

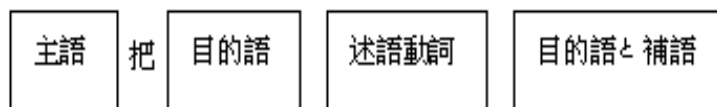


図2 五段モデル

我々は大量の教科書、科学技術文献から「把」文型の例文を収集して構文解析し、「把」文型の構文解析特徴を抽出する。実は五つの部分からなる。したがって、図2五段モデルの構文解析の手法を利用する。

3. 語順の変換

中国語語順は日本語語順と異なるため、語順変換規則を調整しなければならない。翻訳過程で4の語順変換を行う。五段モデルでの中国語日本語の語順変換の例として以下に示す。

中国語：主語 把 目的語 述語動詞 目的語/補語

日本語：主語 目的語 目的語/補語 述語動詞

しかも、正しく語順を決定することは中日機械翻訳システムを構築する際に重要な課題である。まず、把文型構造の特徴により主語、目的語、述語動詞に区切る。ここで問題となるのは、把文型中に1つ以上の動詞が含まれるので、述語動詞を決定することが困難である。そこで、我々は教科書、科学技術文献中の把文型例文を収集し、それらを解析し、述語動詞を決定するルールを作成した。このルールより述語動詞を決定し語順の変換を行う。ルールを表1に示す。

4. 格助詞の確認

日本語の格助詞に対応して中国語文に現れる語を本論文では補助語という。一般に、日本語の一つの格助詞はいくつかの中国語の単語と対応する。この問題を解決する方法の一つは異なる補助語に対応して

異なるコードをつくることである。しかし、教科書、文献、文法書関係の約 12,000 文から格助詞を抽出し、対応する補助語について考察を行った結果、格助詞「で」の補助語候補数は 23 個、格助詞「に」の補助語候補数は 23 個であった。格助詞「で」と対応する構造ルールは表 2 を示す。

表 1 述語動詞の確認ルール

V+的+N	Vは述語動詞ではない
“把”+V	Vは述語動詞ではない
的+V	Vは述語動詞ではない
V+V1(来, 一下…)	V1は述語動詞ではない
V1(帮我, 给我…)+V	V1は述語動詞ではない
センテンス最後+給我	“給我”が述語動詞
V(在, 到)+場所	このVを優先する
V(成)+変身したもの	このVを優先する
V(給)+相手	このVを優先する

格助詞利用の例

例 2. 到四点能把私の行李寄放在这里？ (場 所)

訳文: 四時までここで荷物を預かってもらえますか。

例 3. 請用航空把这封信寄出去。 (手 段)

訳文: この手紙を航空便で出してください。

例 4. 我好疲倦, 请在三小时后再把我叫醒。 (原 因)

訳文: かなりなので、三時間後に起こしてください。

例 5. 可不可以用同样的价钱把六楼的房间换为二十楼？ (お 金)

訳文: 六階の部屋から二十階の部屋に同金額で変更できますか。

表 2 格助詞「で」に対する構造ルール

体 言	補助語
1. 場 所	在
2. 交通手段 (3類)	騎
3. 交通手段 (2類)	乘
4. 交通手段 (1類)	坐
5. 手 段	用
6. 原 因	因
7. お 金	花
など	

5. 時制の確認し日本語述語の時制を決定

センテンスの時制は現在形及び過去形は述語動詞により決定されるため、我々は大量の把文型の例文を解析し現在形過去形の時制ルールを作成し表 2 に示すようにまとめた。未来形は主語により決定されるため、別にルールを作成する必要がある。

我々の時制確認研究では三つの部分を考える。収集したものは以下に示す。

表 3 時制確認研究の三つの部分

1. 時間名詞	前天 (おととい) / 昨天 (昨日) / 今天 (今日) / 明天 (明日) / 後天 (あさって) など
2. 時間副詞	已經 (もう、すでに) / 將要 (まさに) など
3. 助詞	着/了など

詳しいルールは以下のとおりである。

例 6、昨天他已經把花送去了。

訳文： 昨日彼は花を送りました。

1. アスペクト助詞 “了”，“着”

“了” のみで成立する “把” 構文は過去型の「た」文と対応する場合が多い。

“着” のみで成立する “把” 構文は命令文になりやすい。

例 7. 我已經把灯关了。

訳文： 私は電気を消した。

2. 動詞が繰り返して出現する場合，例えば，「包一包」。すべて「た」型を使わない。

例 8. 他把傷口包了包。

訳文： 彼は傷口に布を巻きつける。

3. 動詞の後に補語がある。後に文節が続かない場合，完了した動作を述べるには全て “了” を用いなければならない。

例 9. 他把藥放在了卓子上。

訳文： 彼は薬を机の上においた。

4. 動詞の後に目的語がある。後に文節が続かない場合，完了した動作を述べるには全て “了” を用いなければならない。

例 10. 售貨員把應該找的零錢給了我。

訳文： 店員は返すべきおつりを渡してくれた。

5. 動詞の前に状語がある。この場合の連用修飾語は結果や完了の意味が含まれる。

例 11. 我把他向外拉。

訳文： 私は彼を外に引っぱった。

6. 翻訳実験と考察

我々は本論文で提案した手法による「把」文型機械翻訳実験システムを構築した。現在，実際の「把」文型のセンテンスを入力して，翻訳結果の正確率を検証するところである。実験によって本論文で提案した翻訳手法の正確性と有効性が確認されている。

7. まとめ

本稿では，中日機械翻訳における “把” 文型の処理手法について述べた。私が収集する把字句は28篇のテキスト (1, 158, 248漢字) から無作為に選んでいる。把字句は2086条である。現段階では，実験を通じて本手法の適応性を検証しているところである。今後の課題としては，システムの精度を高め，実用的なシステムの中に組み込む予定である。

【参考文献】

- 1) 范莉馨 任福継 宮永喜一 柄内香次：中日機械翻訳にあける中国語複合語の自動合成について 情報処理学会論文誌, vol. 33 no. 9
- 2) 范莉馨 任福継 宮永喜一 柄内香次：中日機械翻訳における離合詞の処理手法 情報処理学会論文誌 vol. 35 no. 9
- 3) 范莉馨 任福継 宮永喜一 柄内香次：関連語を用いた文の分解に基づく中日機械翻訳システム 情報処理学会論文誌 vol. 35 no. 12