

## 文構成素の適切な選択と配置による自然な日中機械翻訳

白 静

野村 浩郷

九州工業大学情報工学部  
〒820-8502 飯塚市川津 680-4

*nomura@ai.kyutech.ac.jp*

*http://www.dumbo.ai.kyutech.ac.jp/nomura/*

あらまし 日中機械翻訳システムが実用システムとして多数に提供されるようになってきた。しかし、不適切な翻訳や不自然な翻訳が多く見られ、実用システムとして使用するには難がある。本論文は、多数の例文を挙げながら、中国語の文構成素間に固有の制約関係を利用して文構成素配置による、自然な中国語訳文の獲得について述べる。最初に、中国語の翻訳結果からいくつかの修正する必要の文型をまとめる。その文型に対し、具体的な調整方法を述べる。日本語と中国語の名詞化用接続詞および他構文要素接続の相違があり、日本語の形容詞を述語とする同じ文法表現であっても中国語としてさまざまな言語表現もあり、日本語原文には欠けている名詞句情報の認識と補完もある。最後に文表現における使役、受身時制などの法情報について、両言語文要素間の依存関係の相違について述べる。

## Suitable Selection and Proper Disposition of Sentence Constituents for Generating Natural Translation in Japanese-Chinese Machine Translation

Bai Jing

Hirosato Nomura

Kyushu Institute of Technology  
Iizuka 820-8502 Japan  
*nomura@ai.kyutech.ac.jp*

**Abstract** This paper presents the selection and disposition of sentence constituents focused on the generation of natural Chinese translation. We have examined many translation results for simple grammar Japanese sentence using some free online translators and one business application system. We found in order to get a correct and more natural translation it needs not only the constraint information of original language, but also necessary for the target language. We have collected some sentence patterns there is different constraint information between Japanese and Chinese language. Through the analysis of difference to finding out the more suitable selection and proper disposition, then getting more natural translation. Although our sentence patterns is for simple sentence and lacks categories, it gets better results than other translation systems.

## 1. はじめに

我々は、従来より、CLINT (Cross LIngual Information Technology) と呼ぶ一連の機械翻訳の研究を進めてきた。CLINT には、現在、日タイ機械翻訳システム CLINT J2T [1-3]、日マレー機械翻訳システム CLINT J2M [4]、日西機械翻訳システム CLINT J2S “JEMS” [5]、および日中機械翻訳システム CLINT J2C がある。いずれも小規模な実験システムを開発している。

近年、日中機械翻訳システムがいくつか実用システムとして提供されるようになってきた。しかし、不適切な翻訳や不自然な翻訳が多く見られ、実用システムとして使用するには難がある。

本稿は、日本語から自然な中国語訳文を得るための言語変換データを整理することを目的とする。そのため、まず、現在最新公開の日中機械翻訳システムによる翻訳品質の分析・考察に基づいて、その翻訳結果の正しさだけでなく、中国語として自然であるかどうかについて整理する。続いて、中国語の文構成素間に固有の制約関係について細かく発見された問題点を一つ一つ例文を挙げながら述べ、翻訳変換における問題点を整理する。そして、分析により発見された問題点に対して、自然な中国語文を得るために、適切な訳語選択と文構成素配置に関して、具体的な解決方法を提案する。それに基づき、日本語・中国語の小規模な機械翻訳システムを作成する。最後にその機械翻訳システムによる翻訳実験とその考察を行う。なお、上記分析は、文法的網羅性を確保するため、初級レベルのものではあるが、教科書[6]を参照した。

## 2. 誤りの多い文型

現在最新公開の日中機械翻訳システムの考察に基づいて、その翻訳結果の正しさではなく、訳された中国語が自然であるかどうかという点に着目し、それらを具体的に分析・整理して、中国語の文脈制約により発見された問題点について例文を挙げて逐一述べる。

具体的な文型の誤り:

- 指示代名詞と数詞を含む名詞句  
<名詞句> ::= <指示代名詞><名詞>  
<名詞句> ::= <数詞><の><名詞>
- 形容詞と形容動詞を含む名詞句  
<名詞句> ::= <形容詞><名詞>  
<名詞句> ::= <形容動詞><名詞>  
<名詞句> ::= <名詞><の><名詞>
- サ変動詞  
<サ変動詞> ::= <～する> | <～をする>
- 数詞を含む日本文  
<文> ::= …<数詞><動詞> …

- 頻度を含む日本文  
<文> ::= 頻度を含む文
- 述語＝形容詞と形容動詞  
<文> ::= <…は><形容詞>  
<文> ::= <…は><形容動詞>
- 述語＝動詞可能形  
<文> ::= <…は><…が><動詞可能形>  
これから、以上の文型に対して、一つずつ説明していく。

## 3. 中国語における量詞の役割

中国語では、数詞はほとんど単独で使えるが、名詞を直接修飾することができない。“一筆”とは言わない、“一支筆”である。

量詞は事物や動作の量を計る単位である。量詞は単独で使えないだけではなく、名詞を直接修飾することもできない。数詞や指示代名詞あるいは疑問詞と結合ながら使われる。“個学生”とは言わない、必ず「数詞＋量詞」の形で“一個学生”と言わなければならない。

普通は数詞と量詞を組合せて「数量詞」として使う。例えば、“一個”“三包”“五支”など。厳密に言うと、数量詞は品詞ではなく、あくまでも数詞と量詞の連語である。

したがって、中国語文法で言う「数量詞」は真の数量詞ではなく、「数詞＋量詞」と「指示代名詞＋量詞」の連語構造ということになる。

日本語と同じく、数量詞は名詞を修飾場合、文脈によっては、その名詞の代わりとして使われる。例えば:

“我有兩台電腦，一台放在公司，一台放在家裏。”(「僕はコンピューターを2台持っている。1台は会社に置いてあり、1台は家にある。」)

数量詞は名詞を修飾する時、制限がかなりある。すべての名詞に適用できる量詞が存在しない、それぞれの名詞に対して、それぞれの専用な量詞がある。しかし、今までの機械翻訳システムは「指示代名詞＋名詞」「数詞＋の＋名詞」という構造を持つ日本語の名詞句を中国語に翻訳する時、ほとんど“個”という量詞を使って、翻訳を行っている。

この問題について今の汎用な量詞「個」に代わって、一つの名詞に1つの量詞という属性をつける。

日本語:

<名詞句> ::= <指示代名詞><名詞>

中国語:

<名詞句> ::= <指示代名詞><専用量詞><名詞>

日本語:

<名詞句> ::= <数詞><の><名詞>

中国語:

<名詞句> ::= <数詞><専用量詞><名詞>

他の「あれ」「これ」にも量詞がいるときもある:

I、(テレビを指して)これはカラーだ。

○ 台是彩色的。

II、(自転車を指して)それが折りたたみ式だ。

○ 那 是折叠式。

III、(教科書を指して)あれは分りにくい。

○ 本不好懂。

#### 4. 中国語における「的」の役割

<名詞句> ::= <形容詞><名詞>

<名詞句> ::= <形容動詞><名詞>

<名詞句> ::= <名詞><の><名詞>

一番目と二番目の形容詞と形容動詞を含む名詞句の場合には、日本語では形容詞と形容動詞を名詞の直前に置いて名詞を修飾する。中国語では実際にある形容詞を日本語と同じように名詞の直前に置いて名詞を修飾することができるが、有る形容詞は名詞を修飾する時、必ず形容詞の後ろに「的」という助詞を付けて、名詞を修飾する。

三番目は: <名詞句> ::= <名詞><の><名詞> の場合ほどのような「の」を「的」に訳してはいけなだろうか。「的」に訳せない「の」は以下の通りである:

・ 時間詞と時間詞の間には「的」を使わない:

時間の間にある接続語「の」の中国語訳語「的」は要らない、「今日の午後」は直接に訳すれば「今天的下午」であるが、実際は「今天下午」でよいのである。つまり、「の」を訳さないで省略すればいい。他の表現:

来年の3月 → 明年的3月

今日の正午 → 今天中午

昨日の3時 → 昨天3点

特に「朝」「晩」「午前」「午後」「時間」といった単語の前の「の」は「的」と訳さないで、省略していい。

・ 数量詞と名詞の間には「的」を使わない:

「1つの梨」「2本の箸」は「一只的梨」「5 双的筷子」ではなく、「一只梨」「5 双筷子」である。「時間詞」の場合と同じ、訳さないで省略すればいい。他の表現:

1足の靴下 → 一棵

2つの映画 → 两 蛋糕

3枚のシャツ → 三 票

#### 5. サ変動詞と動詞句

<サ行動詞> ::= <~する>|<~をする>

誤った訳文が得られた原因を発見するために、以下の例対について考察した:

日本語	中国語
・会議で発言する	○在会 上 言(自然な訳文)
・会議で発言をする	?在会 上做 言
・太平洋を航海する	○在太平洋航海(自然な訳文)
・太平洋航海をする	?做太平洋航海

例文が示すように、日本語原文では同じ意味の文であるが、助詞「を」を加えるかどうかによって、違う中国語訳文が得られていた。そして、助詞「を」を加えない日本語原文は自然な中国語翻訳結果が得られた。

これは、日本語のサ変動詞の書き方の違いに原因があると考えられる。「する」の直接的な中国訳語は「做」という動詞である。しかし、「する」はサ変動詞として使われる場合、その自然な中国語訳文とは、「する」を無視して、前の名詞だけ、翻訳したらいいのである。したがって、中国語に翻訳するとき、「航海する」の場合、「做航海」ではなく、「航海」だけでよく、「調査する」は「做」ではなく、「」だけでよいのである。上の例の示すとおりである。

その解決方法とは、日本語原文に対して、構文的に書き換えることを行う:

書き換え前:<サ行動詞> ::= <~をする>

↓

書き換え後:<サ行動詞> ::= <~する>

#### 6. 両言語の補語の違う役割

<文> ::= ...<数詞><動詞> ...

この問題を説明する前に、まず、中国語と日本語の文の対応について簡単に説明する必要がある。

日本語原文:

連体修飾語 主語 補語 連用修飾語 目的語 述語

私の 妹は 一年間 中国で 中国語を 学んだ

中国語訳文:

連体修飾語 主語 連用修飾語 述語 補語 目的語

p-4 4 , p-4

日本語の原文「私の妹は中国で一年間中国語をんだ」を中国語に翻訳すると、「一年間」は補語として翻訳された。この補語は目的語「中文」の補語でなくて、述語「学習了」の補語である。中国語では的語の前に置かなければならない。

次の訳文は時間量詞と目的語の位置によって、現れた誤りである：

I、彼は大きい林檎を五つ買った。

? 他 了大的苹果五个。

II、昨日1時間テレビを見た。

? 看了 昨天1小 。

III、紙を1枚持ってきてください。

? 拿来 1 。

IV、彼は昨日サッカーを2時間した。

? 他昨天踢了足球2小 。

V、私は中国語を2年間勉強した。

? 我学 了中文2年。

以上の誤文を直すと以下のようになる：

I、彼は大きい林檎を五つ買った。

○ 他 了五个大苹果。

II、昨日1時間テレビを見た。

○ 昨天我看了一个小 。

III、紙を1枚持ってきてください。

○ 拿来1 。

IV、彼は昨日サッカーを2時間した。

○ 他昨天踢了两个小 足球。

V、私は中国語を2年間勉強した。

○ 我学 了两年 。

## 7. 中国語の「都」の役割

中国語では複数の物や人が同じ状況、同じ動作を行う場合、その状況や動作の前に“都”を使わなければならない。例えば：

I、昨天和今天都 到了。

昨日も今日も遅刻した。

II、中国菜和日本菜都好吃。

コーヒーも紅茶もおいしい。

<文> ::= 頻度を含む文

他に、日本語原文の中には「みな、すべて、共に」がなくても、中国語訳文に“都”を付ける必要がある。具体的にいえば、“每天、毎星期、毎年、毎次”などが含まれる文である：

I、先生は毎年海外旅行へ行きます。

? 老 年去海外旅行。(不自然)

○ 老 年都去海外旅行。(自然)

II、李さんは毎週部屋の掃除をする。

? 小李 周打 房 。(不自然)

○ 小李 周都打 房 。(自然)

以上の例を見て分かるように、定期的なことを強調するとき、“都”を忘れられない。ただし、質問文に“都”は必要がない。例えば、“毎月何回遠足しますか。”(你 月去几次郊游?)で十分で、“你 月都去几次郊游?”とは言えない。

## 8. 同じ文型の違う表現

<文> ::= <…は><形容詞>

日本語で形容詞は述語としてよく単独で使われる、中国語では形容詞も述語として使われるが、ほとんどの場合、程度副詞の修飾が必要です。特定の場合だけ単独で述語として使える：

・ 選択の対比文：文の前後に2つの事情に対して正反対の意味を出し、対比する場合：

I、沙漠 ，山中冷。

(沙漠は暑い、山の中は寒い。)

II、部 和 ，科 。

(部長は優しいが、課長は厳しい)

III、柑橘甜，苦瓜苦。

(蜜柑は甘い、苦瓜は苦い。)

・ 答えの文：質問文に対する答える場合：

I、A: 站 是公园 ?

(駅は遠いですか、それとも公園は遠いですか。)

B: 站 。

(駅は遠いです。)

II、A: 你家 是他家 ?

(お宅が遠いですが、それとも彼の家が遠いのですか。)

B: 我家 。

(私の家のほうが遠いです。)

・ 比喩の文：

I、眼睛像天空一 。

(眼は空のように青い。)

II、皮 像雪一 白。

(肌は雪のように白い。)

対比でもなく、答えでもなく、比喩でもない場合、単にある相手に有る様子や状況について紹介したり、説明したりする場合、形容詞の前に程度副詞をつける。例えば：

I、花は赤いです。

- 花很 。
- II、雪は白いです。
  - 雪很白。
- III、このお茶はおいしいです。
  - 个茶很好喝。

否定の場合は否定副詞の“不”をつければ、程度副詞は必要ない。例えば：

- X、この服は綺麗ではない。
  - 件衣服不漂亮。
- XI、彼らの大学は広くない。
  - 他 学校不大。
- XII、このお茶はおいしくない。
  - 个茶不好喝。

以上のように、形容詞が単独で述語になれるのは、対比文と答えの文だけである。それ以外は副詞“很”、“非常”と一緒に使わなければならない。

## 9. 文型の書き換え

<文>::=<…は>< …が><動詞可能形>

これは日本語の動詞可能形についての誤りである：

日本語	中国語
彼は英語が話せる	? 他英語能 (不自然な文)
彼は英語を話せる	○他能 英語

私は中国語の新聞が読める	? 我中文的 能 (不自然な文)
私は中国語の新聞を読める	○我能 中文的

あなたはお酒が飲めますか	? 你酒能喝 (不自然な文)
あなたはお酒を飲めますか	○你能喝酒

彼は日本語で手紙が書ける	? 他用日 信能写 (不自然な文)
彼は日本語で手紙を書ける	○他用日 能写信

例文を見れば分かるように、左側の二つの日本語原文は違う助詞を使っているが、意味は同じである。しかし翻訳されると、それぞれ違う中国語訳文になった。右側の訳文の示すとおりである。「彼は英語が話せる 彼は英語を話せる」という例文を見てみる。日本文原文は助詞「を」を使った場合、「英語」は目的語として認識され、意味の正しい中国語訳文が得られたが、助詞「が」が使われた場合、「英語」は目的語ではなく、主語として認識された。

これは構文構造を変換する場合に現れた問題である。この問題に対して、まず、文型の書き換えをしてから、翻訳を行う：

<文>::=<…は>< …が><動詞可能形>

↓

<文>::=<…は>< …を><動詞可能形>

## 10. 日本語・中国語の文脈表現上の相違

### 10. 1. 主語の処理：

・補足：

<日本語> A: 田中さん、読んでください。

<中国語> A: [田中], [貴方][読む]

<日本語> B: お母さん、おなかが痛いよ。

<中国語> B: [お母さん], [私][お腹][痛い]

日本語は主語を省略するのが普通であるが、中国語では主語がほとんど省略されない。その原因の一つは中国語の名詞は日本語のように言い分けることをしないからである。例えば、“我媽媽—你媽媽”(母—お母さん)、“我丈夫—你丈夫”(主人—ご主人)。もう一つの原因は中国語の動詞は、日本語の「命令」「勧誘」「謙遜」「丁寧」などのような形の変化がなく、その役割を主語がはたしているのである。例えば、“我做吧—你做吧”(やりましょう—やってください)、“我拿吧—你拿吧”(お持ちしましょう—お持ちになってください)。以上のように、中国語は主語の存在が非常に大切である。

・重複

<日本語> A1: [王さん]はどこにいるのですか。

<中国語> A1: [王さん][いる][どこ]

<日本語> A2: 教室です。

<中国語> A2: [彼][いる][教室]。

<日本語> B1: [母]は賛成しないよ。

<中国語> B1: [私母]ない賛成

<日本語> B2: もっと[お母さん]を説得したら。

<中国語> B2: [貴方][もっと][説得][彼女]

日本語では「あなた」「君」「彼」「彼女」などの人称代名詞があまり使われない、特に「彼」「彼女」は特別の場合を除いてほとんど使わない。また、日本語では、何度繰り返す場合でも、その都度名前や身分を言うことが多いが、中国語では名前が一度出れば、

その後の文に代名詞を使えばよい。目上の人に対しても同じである。

## 10. 2. 法情報:

日本語の文構造は、格文法で動詞について色々法情報が付加される。それらは、相情報、モード情報、および時制情報からなる。

法情報についての日本語の例文と中国語の対応関係を見てみる:

- ・ 相情報:(使役と受身)

日本語では、可能、受身、使役、時制などの法情報を表現する時、助動詞「できる」「される」「させる」などの助動詞を用いて表す。中国語では日本語のように動詞の活用がないが、これらの法情報を表現するために、動詞の前後に、または、文末に異なる「単語」を加えて表す。

<日本語> A: 家が[焼かれたた]。

<中国語> A: [家][された][焼く]

<日本語> B: 僕の自転車が[盗まれた]。

<中国語> B: [僕の自転車][された][盗む]

<日本語> C: 父は僕に酒を買いに[行かせた]。

<中国語> C: [父][させる][僕][行く][買う][酒]

日本語の相情報「される」、「させる」を中国語に訳する場合:動詞の前に相当する中国語の「される」「させる」を置。

- ・ モード情報(てみる、つもりだ)

<日本語> <中国語>

食べてみる。

<動詞>+てみる → <動詞>+てみる

明日行くつもりだ。

<動詞>+つもりだ → つもり+<動詞>

日本語のモード情報「てみる」を中国語に訳す場合:動詞の後ろに相当する中国語の「てみる」を置く、「つもりだ」の場合:動詞の前に相当する中国語の「つもり」を置き換える。

- ・ 時制情報(た)

<日本語> <中国語>

A1: 昨日何を買った。 → 昨日 買う た 何

A2: カメラを買った。 → 買う た カメラ

<動詞の過去形> → <動詞>+た

日本語の時制情報「た」は動詞の過去形である。中国語に訳すとき、動詞の後ろに過去形を表す単語を置く。

## 11. むすび

前述した解決方法に基づき、条件付きの翻訳パターンを作成し、日本語から中国語への実験システムを構築した。そのシステムは中国語の文構成素間にある制約関係を利用して、簡単な例文を解析することが可能になった。

これから、本論文で述べた文構成素の制約関係のみならず、文脈的な制約も含め、他の表示レベルと変換規則を整備し、それに基づいての言語モデルとシステムモデルの構築は今後の課題となる。

## 参考文献

- 1] Chumphol Krootkaew, Hidetoshi Nagai, Teigo Nakamura, Hirosato Nomura: “Towards Machine Translation using Cross-language Information”, Proc. of International Conference on Machine Translation & Computer Language Information Processing, pp80-84, 1999
- 2] Chumphol Krootkaew, Hidetoshi Nagai, Teigo Nakamura, Hirosato Nomura: “Transferring Patterns Based Japanese to Thai Machine Translation”, AAMT Journal, No.28, pp19-28, 2000
- 3] Hanani bt. Abdul Wahab: “Constrastive Study of Japanese and Malay Languages and its Application to Machine Translation”, 平成 12 年度卒業論文、九州工業大学、2001
- 4] Manuel Medina Gonzalez and Hirosato Nomura: “A Cross-Lingual Grammar Model and its Application to Japanese-Spanish Machine translation”, 情報処理学会研究報告、Vol.2006、No.94、自然言語処理研究会、2006-NL-175、pp119-126, 2006
- 5] 野村 浩郷:“自然言語処理の基礎技術”, 電子情報通信学会, 1988。
- 6] 大内 田三郎:基礎からよくわかる中国語文法参考書、駿河台出版社、2000
- 7] 士文:“代 法信息 典 解”、中国清化大学出版社”、2003
- 8] 郭春貴:“誤用から学ぶ中国語”、白帝社、2001