

TSPO特徴を利用した中日機械翻訳

任 福継

広島市立大学情報科学部知能情報システム工学科

本稿では、中国語の複文分解及び各システムを検討し、動詞の細分類を行い、中国語が主題型言語であるという観点から、中日両言語の特徴を抽出し、TSPO性質の中日機械翻訳システムへの応用手法を提案する。ここで、Tは文の主題 (Theme)、Sは文の主語 (Subject)、Pは文の述語 (Predication)、Oは文の目的語 (Object) を示す。中国語文法の深層構造にはTSPO四つの項があり、TはS、P、Oと同一のレベルで使われている文法項である。実際にはこの深層構造TSPOからTSP、TPO、TSPOという表層構造に派生している。一方、日本語も主題型言語であると考えられる。その深層構造はTSOPであり、TSOPからいろいろな表層構造に派生している。この観点を用いると、従来の文法理論では解決できなかったいくつかの中国語特殊な文型は正確に解釈でき、正しい日本語訳文を生成できる。

Chinese Japanese Machine Translation System Using the theme-type characteristics

Ren Fuji

Faculty of Information Sciences, Hiroshima City University
ren@its.hiroshima-cu.ac.jp

Abstract This paper tries to examine the theme-type characteristics of Chinese and Japanese, and presents a new method for Chinese-Japanese Machine Translation (CJMT) using the theme-type characteristics. We consider that Chinese and Japanese are theme-type languages, whose basic syntactic structures are Theme-Subject-Predicate-Object and Theme-Subject-Object-Predicate, respectively. This paper discusses the transition relationships between the deep structure and the surface structures of the syntax.

1. はじめに

機械翻訳に関する研究における主な問題としては構文解析と意味解析があると考えられる。数十年の努力により構文解析は既に完成され、残ったのは意味解析であると言われているが、この観点は印欧語族を源言語とする機械翻訳に対して適当であるかもしれないが、日本語と中国語などを源言語とする機械翻訳に対してはあまり妥当でない。特に、中国語の構文解析を行う際に、英語の解析に使っている文法理論をそのまま流用すると大きな問題が生じる。

著者らは中日両言語の機械翻訳システムを研究開発しているが、英日と英中間の機械翻訳と比較すると、未開拓の部分が極めて多い。以下に示すような中国語の特徴により、中国語の構文解析が非常に難しいと考えられる。

(1)中国語では、単語と単語の間には、単語を区別できるスペースがなく、日本語のように助詞、助動詞などの機能語を平仮名で表して構文要素を区切ることもない。

(2)中国語では、動詞、助動詞、形容詞などの語尾変化、すなわち活用現象もない。

(3)中国語では、兼用品詞（二つ以上の属性をもつ単語）が多い、また、兼用品詞は形態上の区別がない。

(4)中国語文法において、英語のような主語、述語の形式化定義はとて難である。

(5)中国語文の実際上の使用度から見れば、単文は少なく、複文の方が多い。

我々は大量の中国語複文の調査結果から、複文を構成する各単文の間には一般に特定の「接続要素」（関連語という）が存在することを見出した。さらに、中国語複文と日本語文との対応関係を検討し、関連語を用いて中国語複文をいくつかの独立な単文に分解し、関連語の性質を用い、各単文の翻訳結果から最終の訳文を生成する手法を提案した。実験により、この手法は複文の分解に対してはとて有効であるが、残ったのは単文自身の解析である。例えば、動詞「吃」に対して、以下の例がある。

- 例1 吃苹果 (りんごを食べる)
- 例2 吃大碗 (大きな皿で食べる)
- 例3 吃館子 (料理屋で食べる)
- 例4 吃敗戰 (戦いに負ける)
-
- 例5 張三吃了 (張三は食べた)
- 例6 生魚片吃了 (刺身を食べた)
- 例7 鷄吃了 (?)

このような問題の解決については、動詞の属性及び動詞と体言との関係、すなわち格関係を明らかにしなければならない。本稿では、中国語の動詞格システムを検討し、動詞の分類を行い、さらに、中国語が主題型言語であるという観点から、中日両言語の特徴により、TSPO性質の中日機械翻訳システムへの応用手法を提案する。ここで、Tは文の主題

(Theme)、Sは文の主語 (Subject)、Pは文の述語 (Predication)、Oは文の目的語 (Object) を示す。

中国語が主題型言語である観点からすると、中国語文法の深層構造にはTSPO四つの項があり、TはS、P、Oと同一のレベルで使われている文法項である。実際にはこの深層構造TSPOからTSP、TPO、TSPOという表層構造に派生している。一方、日本語も主題型言語であると考えられる。その深層構造はTSOPであり、TSOPからいろいろな表層構造に派生している。この観点をを用いると、従来の文法理論では解決できなかったいくつかの中国語特殊な文型は正確に解釈でき、正しい日本語訳文を生成できると考える。

以下では、第2章で中国語複文の分解、第3章で中国語の特徴特に語義依存性、第4章で中国語の格システム、第5章で中日両言語のTSPO特徴、第6章でTSPO特徴を中日機械翻訳システムへの応用について述べる。

2. 中国語の複文分解

複文の例 (例8) : 因為天氣不好, 所以我們沒去頤和園。

日本語訳文: 天氣が良くなかったので、私たちは頤和園に行きませんでした。

我々の定義によって、上の例では、

関連語は「因為.....所以」であり、

単文1は「天氣不好」であり、

単文2は「我們沒去頤和園」である。

関連語の共起関係により一語性と多語性に分けられる。詳しくはここで省略するが、以下に複文の分解アルゴリズムの概要及び複文の種類を示す。

関連語を抽出するために、関連語テーブル表1を設けている。表1中、“キーワード1” (前) は文の分解及び関連語の判断に使用される情報である。一語性関連語についてはそれ自身であり、多語性関連語については前半部分である。“呼応ワード2”

(後) は多語性関連語の後半部分である。“MC” は1つの関連語が複数の日本語表現と対応する場合に使用される訳文条件テーブルへのポインタである。

また、「複文分解の条件」には文の分解の条件を記録している。例えば、例8の「因為……所以」は関連語テーブルのキーワード1欄と呼応ワード2欄に登録しているので、照合が成功する。さらに、複文分解の条件と照合し、勿論、本例では、成功する（条件の詳細は省略した）ので、関連語「因為……所以」を抽出すると同時に、二つの単文に分ける。全体の合成は訳文関数により行われる。

表2は中国語複文の例である。

表2 中国語複文の種類

NC	複文類型	例文	意味
01	並列複文	人数既多、意見又不一致。	人数が多いし、意見が一致しない。
02	対比複文	他每天不是看书、就是睡觉。	彼は毎日本を読んでいなければ、寝ている。
03	承接複文	我一收到来款、就给他回信。	私はお金を受け取ると、すぐ彼に手紙で返事をする。
04	累加複文	我不但取宝、而且作科学考察。	私は貴重な経験を吸収するだけでなく、科学的に視察する。
05	選択複文	或者打个全胜、或者打个全败。	完全に勝つか、或いは完全に負けるかである。
06	因果複文	因為天气不好、所以我沒去杭州。	天気がよくなかったので、それで私は杭州へ行かなかった。
07	転折複文	他 然後悔、但是無法挽就。	彼は後悔しているが、しかし袖救済する方法がない。
08	条件複文	只有掌握了日語、才能較好地研究日本文学。	日本語をマスターしておいてこそ日本文学をよく研究することができる。
09	無条件複文	民族性無論好壞、改變都是不容易的。	民族性は良い悪いを問わず、それを改めるのが容易ではない。
10	仮定複文	假如我手里有一部電子計算機、就不会算錯。	もしも私の手元に電子計算機が一台あったら、計算は違いするはずがない。
11	譲歩複文	既使意見通不過、也不該放棄它。	たとえ意見が通らなくても、それを放棄すべきではない。
12	取舍複文	与其匆忙交券、不如多花点時間修改一下。	そそくさと急いで答案を出すより、少し余計に時間をかけて手直してみた方がいい。
13	目的複文	為了不影響工作、他近来減少了社会活動。	仕事に影響をしないために、彼は最近社交活動を減少した。
14	時間複文	自從回国以後、身体就好起来了。	帰国してから身体がよくなりはじめました。
15	連鎖複文	構成複文的單文越多、複文的結構越複雜。	複文を構成する単文が多ければ多いほど複文の構造はますます複雑になる。

表1 関連語テーブル情報（一部）

関連語		複文分解の条件	日本語表現	訳文関数	MC
前	後			類別	
不僅	而且	不僅A、而且B	……だけでなく	<3>	
因為	所以	因為A、所以B	……ので	<2>	
才		—	—	20	
……					

3. 中国語の語義依存性

一般に、中国語単文の構造を図1のように表示する。

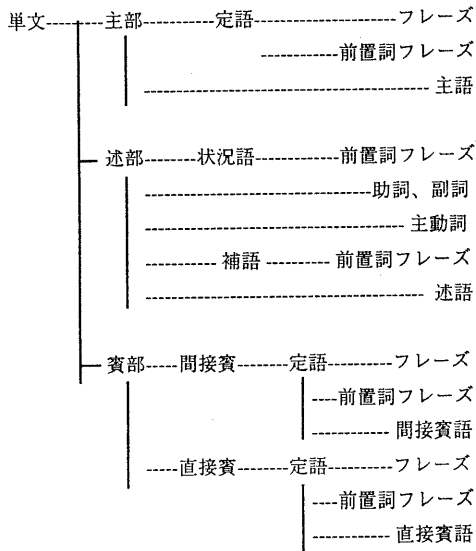


図1 中国語単文の構造

しかしながら、実際にはこのような規則的に並んでない文が多い。

以下に、中国語の機械翻訳における主な特徴について述べる。

(1) 中国語の述語動詞と賓語(目的語)の関係は日本語に訳した場合、述語動詞と目的語との関係になるのは一部に過ぎず、あとは主語と述語になったり、連用修飾語と述語とのさまざまな関係になったりする。つまり、中日両言語では文の成分が必ずしも対応しないのである。

a. 写信 (手紙を書く)

吃饭 (ご飯を食べる)

などは、動作とそれを受けるものとの関係で、日本語でも動詞と目的語との関係になる。

b. 上学 (学校に行く)

去東京 (東京に行く)

などは、動作とその目的地を表わす。

c. 座汽車 (車に乗る)

住飯店 (ホテルに泊る)

などは動作とその行われる場所を表わす。

d. 変良田 (よい田畑となる)

成泡沫 (泡になる)

などは変化とその結果を表わす。

5. 洗冷水 (冷水で洗う)

吃大碗 (どんぶりで食べる)

などは動作とその道具を表わす。

6. 開花児 (花が咲く)

有意義 (意義がある)

などは主語と述語の関係を表わす。

ここから分かるように、中国語の「動詞+目的語」は日本語訳では必ずしも「目的語+動詞」とならない。

(2) 中国語の「動詞+目的語」が表わすさまざまな関係は、二つの単語の意味を離れて形だけで判断することは不可能である。その他の沢山の形も類似である。中国語は孤立語で、「動詞+目的語」の関係も語順だけに頼って表わされ、助詞も使われていないし、語尾変化もないから、述語動詞と目的語の間にどのような具体的意味関係があるのかは、組み合わせられた二つの単語がそれぞれどのような意味をもっているかによって判断しなければならない。たとえ動詞に同一の語を使っても、あとにくる名詞の意味如何によって、両者間に種々の関係が生じる。次にその例をあげる。

<1> 装大米 (お米を入れる)

<2> 装貨 (荷を積む)

<3> 装箱子 (箱に入れる)

<4> 装船 (船積する)

<5> 装幌子 (うわべをかざる)

<6> 装蒜 (とぼける)

以上のように、中国語では語と語の具体的意味関係は、語順以外に語義に頼るところが非常に大きい。以下の章では、この語義関係を明らかにするために、動詞の格システムを検討する。

4. 中国語の格システム

4.1 動詞「打」の意味分類

まず、ある具体的な動詞の意味を検討する。多数の文献と辞書を参考して、動詞「打」の意味は大別して以下ようになる。

<1> 打DA(1) 打つ、殴る。

例：打人 (人を殴る)

<2> 打DA(2) たたく。

例：打鼓 (太鼓をたたく)

<3> 打DA(3) 戦う、攻める。

例：打戦 (戦争をする)

<4> 打DA(4) 壊す、壊れる。

例：碗打丁（お碗が壊れた）

<5>打DA(5) 打って造る

例：打刀（刀を造る）

<6>打DA(6) 元来は手でするある種の動作であったもの。

例：打気（空気を入れる）

<7>打DA(7) かきまぜたりして手をかけて作る。

例：打面包（パンを焼く）

<8>打DA(8) 送る、出す。

例：打電話（電話をかける）

<9>打DA(9) あける、切り開く。

例：打井（井戸を掘る）

<10>打DA(10) えかく。

例：打格子（枠を描く）

<11>打DA(11) 揚げる。

例：打旗子（旗を揚げる）

<12>打DA(12) 禽獣を捕える。

例：打魚（魚をとる）

<13>打DA(13)（案などを）立てる、定める。

例：打主意（方針を立てる）

<14>打DA(14) 従事する、担当する。

例：打雑兎（雑役をする）

<15>打DA(15) ある種のゲームをする。

例：打撲克（トランプをやる）

<16>打DA(16) 種々の身体の動作を表わす。

例：打手勢（てまねをする）

<17>打DA(17) 買う（主として液体のものを）。

例：打酒（酒を買う）

<18>打DA(18) 受け持つ。

例：男子打外、女子打内（男は外を受け持ち、女は内を受け持つ）

<19>打DA(19) 編む、結う。

例：打毛衣（セーターを編む）

<20>打DA(20) 築く。

例：打堤（堤防を築く）

<21>打DA(21) 書類などをつくる。

例：打路条（通行証を出す）

<22>打DA(22) 取り除く。

例：打皮（皮をむく）

<23>打DA(23) 汲む、すくう。

例：打水（水を汲む）

<24>打DA(24) 話す、しゃべる。

例：打招呼（挨拶をする）

<25>打DA(25) 収穫する。

例：打了400斤（400斤収穫があった）

4.2 中国語の格システム

(1) 動詞の細分類

動詞および動詞の表示する動作、状態と関連している主体と客体との間の意味関係により、中国語動詞を以下の6つの種類に分けられる。

<1> 他動詞：主体は動作の動作主であり、動作は客体にかかわる。例：「吃、研究」。

<2> 自動詞：主体は動作の動作主であり、動作は客体にかかわらない。例：「走、工作」。

<3> 外動詞：主体は動作の動作主でなく、動作は客体にかかわる。例：「知道、失去」。

<4> 内動詞：主体は動作の動作主でなく、動作は客体にかかわらない。例：「病、死」。

<5> 領属動詞：領属関係を表わす動詞。例：「有、具有」。

<6> 系属動詞：系属関係を表わす動詞。例：「是、等於」。

以上のように動詞を細分類して、意味関係の表示が容易になり、各動詞の特徴を描くことができる考える。

(2) 格システム

以下に格システムの24格項目だけを示す。

<A>主題

<1>主題

主体

<2>施事

<3>当事

<4>領事

<C>客体

<5>受事

<6>客事

<7>結果

<D>隣体

<8>与事

<9>同事

<10>基準

<11>伴随

<E>系体

<12>系事

<13>分事

<14>数量

<F>凭借

<15>道具

<16>材料

<17>方式

<G>環境

<18>範囲

<19>時間

<20>空間

<21>方位

<H><理由

<22>依拠

<23>原因

<24>目的

4.3 動詞の格フレーム例

例えば、動詞「吃CHI」は5種類の意味に分けられ、その一例を以下に示す。

吃(1) <他動詞> 意味解釈 (略)

基本枠 [施事(孩子,老王,鷄)+吃+受事(飯,菜,酸的)]

例文：山田吃了生魚片 (山田は刺身を食べた)

山田把生魚片吃了

展開枠

[系事] 山田把生魚片<当早餐>吃了

<朝食として>

[与事] 鈴木<替荒木>吃生魚片

[同事] <除了荒木>大家都吃生魚片

<荒木を除く>

[結果] 這頓飯吃了鈴木<一嘴油>

[基準] 荒木<比我>吃得多

[数量] <一次>他吃了一只鷄

[道具] 中国人<用箸子>吃飯

[材料] <用30元錢>他吃了一斤生魚片

[方式] 他們不吃<包食>了

[依拠] <按醫生說的>他吃了兩片藥

[原因] <因為這腥味>他吃不下去了

[目的] <為了身體>他吃了這菜

[時間] <我進來時>他正在吃

[空間] <在札幌>他吃了生魚片

5. 中日言語のTSPO特徴

5.1 中国語のTSPO

ここに、TSPOに対して、以下の制約を加える。

主題Tは文の話題であること。

主語Sは文の説明部分で、動作主で、動詞の前に位置すること。

述語Pは文の説明部分で、文中の動作過程を示す部分であること。

目的語Oは文の説明部分で、文の授事で、動詞の後に位置すること。

以下の文では主題以外の部分を評言という。

中国語文の深層構造 (図3) の仮定から、図4のような構造木が推論される。

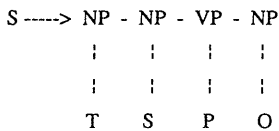


図3 中国語の深層構造

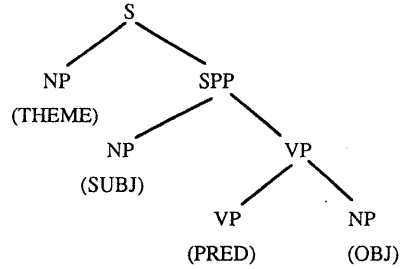


図4 中国語の構造木の例

以下に虚項という概念を導入する。

虚項とは言語構造ではある特徴を持つが、実際の言語現象では出現しない項である。すなわち、虚項は省略による現象である。また、省略には二つの種類がある。一つは文脈による省略、例えば、話題の情景、共通の知識などであり、これを談話省略という。もう一つは同一文中に完全に同じ二つの項が存在するのでその一つを省略することであり、これを文法省略という。

例文(1)は談話省略の例で、省略されたNP1は太郎で、NP2は「這部電影」(この映画)である。

例文(1)：太郎看過這部電影？

[NP1] 看過 [NP2]。

(太郎はこの映画を見ましたか?)

([太郎は][この映画を]見ました。)

例文(2)は文法省略の例である。

例文(2)：那本書他拿走了[NP1]。

(あの本は彼が[あの本を]持ってきた。)

ところが、中国語では主語と目的語も虚項できる。それで、深層構造と表層構造との間に変換規則がある。

次にTPO構造、TSP構造とTSPO構造との関係を示す。

(1) TPO

例文(3)：太郎愛上了花子 <1>

(太郎は花子を愛している)

例文(3)の深層構造は

太郎太郎愛上了花子 <2>

である。

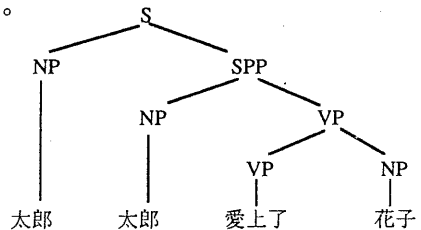


図5 例文3の深層構造

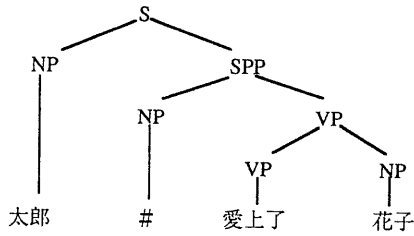


図6 変換過程

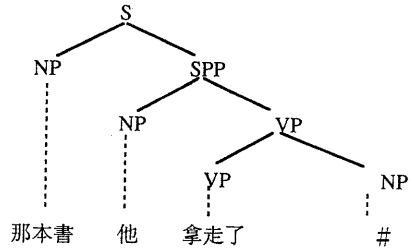


図9 変換過程

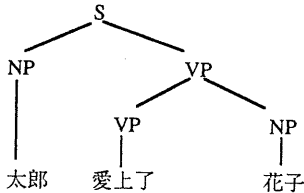


図7 例文3の表層構造

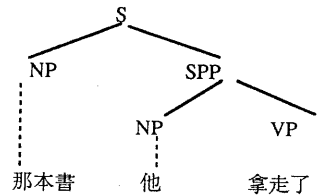


図10 例文4の表層構造

<2>から<1>への変換過程は図5から図7までに示す。

図5から図6までは、同型規則により主語が省略される。

同型規則：TとSが同型なら、Sが省略される。

TとOが同型なら、Oが省略される。

図6から図7までは、動詞フレーズが主述フレーズに代わる。

(2) TSP

例文(4)：那本書他拿走了 <3>

この深層構造は

那本書他拿走了那本書 <4>

である。

<4>から<3>への変換過程は図8から図10までに示す。

図9から図10までは、動詞フレーズが述賓(目的語)フレーズに代わる。

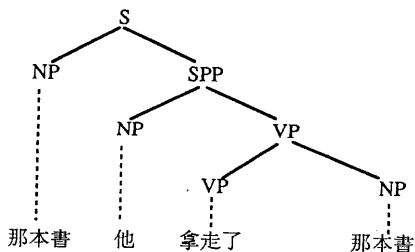


図8 例文4の深層構造

5.2 日本語のTSPO

日本語の主題(特に主題文)について多くの観点から論じられているが、主題に関する定義はあまり明確でないと考えられる。従来、主題に関する議論は主語に対するものが多いが、著者らは、主題は主語だけではなく、その他の成分にも対応して考えなければならないと主張している。すなわち、日本語も中国語と同じ主題型言語であると考えられる。

例文5 太郎は この荷物を 運んだ。

動作者 対象
主題 評言

例文6 この荷物は 太郎が 運んだ。

対象 動作者
主題 評言

上記の例文が示すように、動作者は主題を示す場合もあるし、逆に、評言を示す場合もある。対象も同様である。その他の意味役割も同様である。

統計を取れば、主題になるのは動作者である場合が最も多いであるが、いつもそうなるとは限らない。

従来の研究では主題という概念を深層構造に分離するものが多いが、本稿では、日本語は主題型言語を仮定し、その深層構造は図11のようになると考える。その対応する構造木を図12に示す。ここで、理解の便利のため、主語、目的語、述語だけを用いるが、その他の成分にも同様である。

以下に深層構造と表層構造との間に変換関係を述べる。例文5の深層構造は図13に示す。

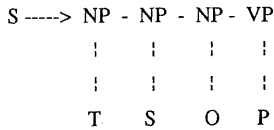


図 1 1 日本語の深層構造

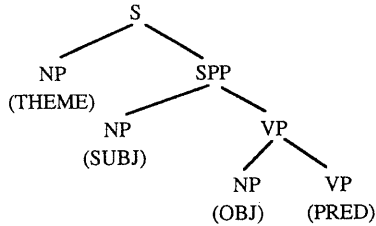


図 1 2 日本語の構造木の例

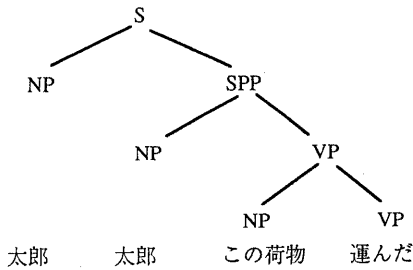


図 1 3 例文 5-1 の深層構造

る。このとき、対応している体言（名詞）の意味属性が必要となる。すなわち、「王太郎」の意味属性と「生魚片」の意味属性に従って、主題が代表している格項目を確認する。例 8 と例 9 の深層構造は図 8、9 であり、これにより、対応する日本語が容易に生成できる。

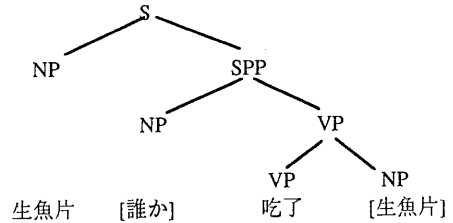


図 1 4 例文 8 の深層構造

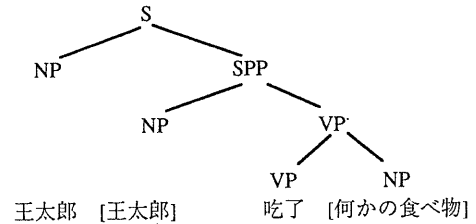


図 1 5 例文 9 の深層構造

6. 中日機械翻訳への応用

中日機械翻訳システムを開発するとき、以下の 2 点を考慮しなければならない。

(1) 主題=主語

この場合に、中国語と日本語とも主語が省略されたが、中国語には、日本語の「は」のように主題を示す形式標識がないので、一般的な文パターンにおいては主題/主語の統語的一体化が見られる。しかし、主題と主語が共起する文において、中国語と日本語と同様に平行的な主題構文があると考えられる。例文 7 はこのような例である。

例文 7: 大象鼻子長 (中国語)
 象は鼻が長い (日本語)

(2) 主題=非主語

この場合に意味的な深層構造を明らかにしなければ、中国語の分析ができなく、正しい日本語訳文を生成できない。

例文 8 生魚片吃了 (刺身を食べた/刺身が食べられた)

例文 9 王太郎吃了 (王太郎が/は食べた)

TSPO を用い、上述の問題を解決できる。まず、動詞「吃」の格フレームにより、その格項目を分析す

7. おわりに

日本語には色々な格助詞があって、それが体言の後ろについて、そのついた語が他の語に対してどんな資格関係（つまり格）に立つかが明示される。しかし、中国語はそういう格助詞が全然なくて、語順が極めて重要である。また、語と語の具体的な関係は語順以外に語義に頼るところが非常に大きい。本稿では中国語動詞の細分類を行い、格システムを検討した。また、中日両言語の TSPO の特徴を検討して、この特徴を中日機械翻訳への応用について述べた。今後、中日両言語の TSPO の対応関係および変換手法を検討し、主要な動詞の格フレームを構築し、さらにこの手法に基づく中日機械翻訳システムを開発すると考える。

参考文献

- [1] 崎山理編、日本語と日本語教育、明治書院、1989。
- [2] 陳力為主編、計算言語学研究と応用、北京言語学院出版社、1993。
- [3] 任福繼、中日言語の TSPO 特徴及び CJMT への応用について、人工知能学会全国大会論文集、pp.637-640, 1994。
- [4] 林光等、動詞大辞典、中国物資出版社、1993。
- [5] 范莉馨、他、文の分解に基づく中日機械翻訳システム、信学技報、NLC93-37, pp.49-56。