

日中機械翻訳システム jaw/Chinese における翻訳実験

— 日本語の文字面を用いた表現や接続表現の翻訳事例 —

穆貴彬、浅井良信、王軼謳、卜朝暉、池田尚志

岐阜大学工学部

1. はじめに

我々の研究室では日本語から中国語へのパターン変換型機械翻訳システム jaw/Chinese の開発を行っている。日英対訳コーパス[1]の例文を用いて、jaw/Chinese の検証・改善をしている。

日本語では用言に後接する機能語を介して2つの文を接続するが（我々はこれを接続表現と呼ぶ）、中国語に翻訳すると三種類の文になる。また、日本語のより固定的な単語の並びや慣用句に対応する中国語は特別な訳語になり、我々は文字面の自立語条件を使用した日本語表現パターンを用いることで翻訳している。本論文では文字面の表現や接続表現の処理方法を考案し、jaw/Chinese に実装し翻訳実験した。いくつかの具体例を示して報告する。

2. jaw/Chinese の概要

jaw/Chinese はパターン変換型の方式で、日本語文の表現パターンに対して中国語の表現構造を対応させるパターン翻訳知識ベースと、中国語の表現構造から文を作り出す線状化関数を軸として構成されている。この他体言後接機能語（取り立て表現）や用言後接機能語に関する翻訳処理を行う部分もある[2]。

中国語の表現構造は表現要素をメンバとして持つオブジェクトからなる木構造であり、C++を用いて実装している。翻訳規則は、日本語の表現パターンに対応する中国語の表現構造を作り出すルールであり、データベース化されている。それを jaw/Chinese が使用できる C++の関数に変換し DLL として管理されている。中国語は中国語の表現構造のメンバを

編集して線状化する関数（オブジェクトのメンバ関数として実装してある）によってルートから再帰的に呼び出して生成される。このように中国語の表現構造を介して翻訳を行う点が jaw/Chinese の特徴である。

接続助詞のような自立語の文節同士をつなぎ合わせる機能語（接続表現を含む）に対して、jaw/Chinese は Addition_FW のパターンを用意している[2]。また、パターン内の文節における制約条件には jaw/Chinese は自立語に単語のグループ化による意味属性だけではなく、特定の単語を指定する制約条件（我々は文字面条件と呼ぶ）も実現している。

3. jaw/Chinese における接続表現の処理

日本語の接続表現がある文は中国語の連詞文になる場合と、補語文になる場合と、状語文になる場合の三種類がある。例えば、「V1 ので V2」は連詞文「因为 V1 所以 V2」に、「V1 くらい V2」は補語文「V2 得 V1」に、「V1 (た) まま V2」は状語文「就那样 V1 地 V2」になる。つまり、日本語の「V1+接続表現+V2」の重文は中国語では三種類の文が対応する。

表1 「ので」の Addition_FW パターン

Type	BID	KID	Semantic	KW	Semantic
	1	2	用言	ので	
fw	2	0			用言

jaw/Chinese における接続表現には機能語をキーワードとし、制約条件をキーワード文節の自立語と

キーワード文節が修飾する文節を持つ Addition_FW パターンを使用する[2]。例として、「ので」のパターンを表1に示す。

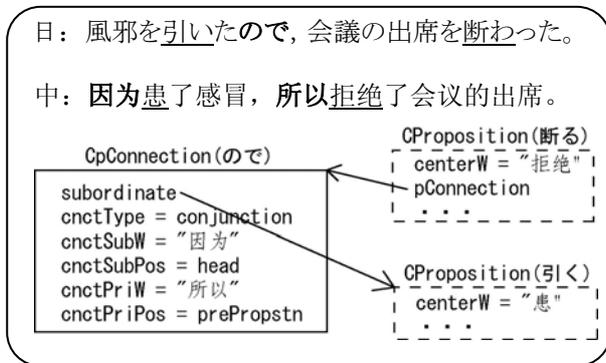


図1 「<用言1>の<用言2>」の中国語の表現構造

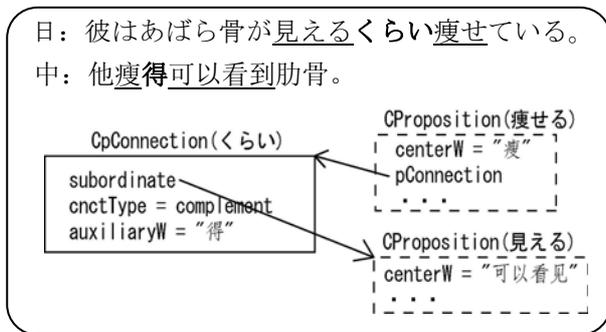


図2 「<用言1>くらい<用言2>」の中国語の表現構造

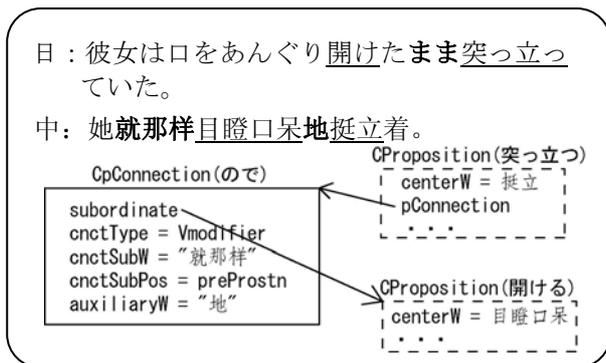


図3 「<用言1> (た) まま<用言2>」の中国語の表現構造

図1、2、3の例文の接続表現は、それぞれ「ので」、「くらい」、「まま」をキーワードとする日本語パターンの翻訳規則とそれぞれの用言1と用言2の翻訳規則によって図1、2、3の表現構造に変換される(簡

略化して示す)。

いずれも接続表現のクラスCpConnectionのオブジェクトを介する構造であるが、どの文型に訳すか、どのメンバに情報を与えるかなどの点で異なる。この構造において、主節(用言2)の線状化関数がメンバpConnectionを介しクラスCpConnectionの線状化関数がメンバsubordinateを介し従属節(用言1)の線状化関数を呼び出し、従属節を線状化するというふうに訳文が再帰的に生成される。連詞文になる場合(cnctType = conjunction)は従属節の訳語を主節の最初に線状化し、連詞はcnctPriPosかcnctSubPosにより主節か従属節の訳語の中に入れる。補語文になる場合(cnctType = complement)は主節の訳語の後ろに補語の助詞「得」入れてから、従属節を線状化する。状語文になる場合(cnctType = Vmodifier)は従属節が主節の訳語の前に線状化し、その訳語の後ろに状語用の助詞「地」を入れることも指定できる。

4. jaw/Chineseにおける文字面パターンの処理

jaw/Chineseは、キーワードにかかる文節には、単語のグループを持つ共通な意味内容を指定する条件としての意味属性だけではなく、特定の単語を単独で用いる文字面条件も使用できる[2]。これを我々は文字面パターンと呼ぶ。文字面の条件とキーワードを組み合わせて、より精度の高い翻訳規則を記述することができる。

文字面パターンは主に以下の三つの翻訳に使用される。

4.1 キーワードの訳が変わる場合

特定の単語と結合するとき、キーワードはふだんの意味を持たず、当単語にあわせて特別な訳になることは、日中翻訳においてよくある。例えば、「立つ」は普段「站立(stand)」の意味を持つが、「<主体>が選挙に立つ」という日本語表現の場合では「参加する(participate)」の意味をとる。そこで、「選挙」

を文字面条件にし、「立つ」をキーワードにし、文字面パターンを用いて翻訳する。そのパターンと中国語の表現構造は図4に示す。

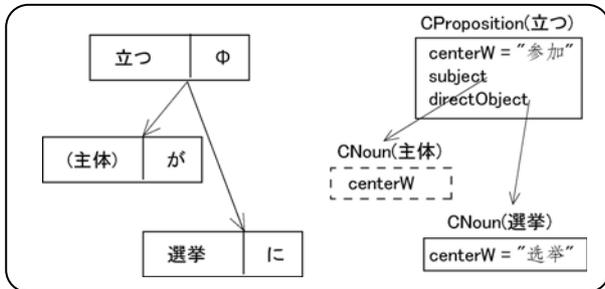


図4 「<主体>が選挙に立つ」

この日本語表現パターンは<主体>に対しては中国語のETにCNoun<主語>が、「選挙」に対してはCNoun「选举」が、「立つ」に対してはCProposition「参加」がある。このETを線状化すると「<主語>参加选举」になる。この例では、日中両言語の構造を比べると、名詞が名詞に、動詞が動詞にきちんと対応しており、構造的にズレがない。もし文字面ノードの「選挙」にかかる修飾要素があれば、中国側では「选举」の修飾要素に変換され、「选举」の線状化関数に呼び出されてそれを翻訳する。

4.2 文字面単語の訳が変わる場合

4.1で述べたのと逆に、キーワードではなく、文字面条件の単語は本来の意味を持たず、特別な訳になることも、日中翻訳においてよくある現象である。例えば、「粘り」は普段「粘性 (stickiness)」の意味を持つが、「<主体>は<物事>に粘りがある」という日本語表現の場合では「毅力(persistent)」の意味を持つ。「粘り」を文字面条件にし、「ある」をキーワードにし、文字面パターンを用いて翻訳する。そのパターンと中国語の表現構造を図5に示す。

この日本語表現パターンは、「主体」に対しては中国語のETにCNoun<主語>が、<物事>に対して対象を現す状語格が、「粘り」に対してはCNoun「毅力」が、「ある」に対してはCProposition「有」がある。このETを線状化すると「<主語>对<物事>有

毅力」になる。

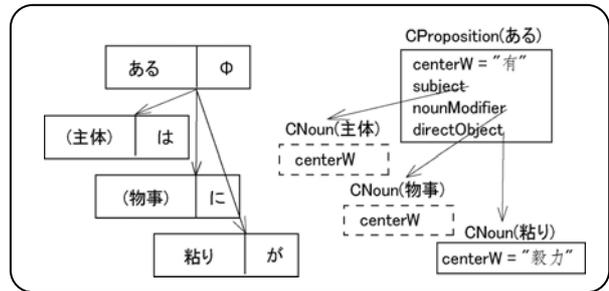


図5 「<主体>は<物事>に粘りがある」

この場合も構造的にはズレがなく、文字面ノードの「粘り」にかかる修飾要素があれば、中国側では「毅力」の修飾要素に変換され、「毅力」の線状化関数に呼び出されて翻訳される。

4.3 構造が変わる場合

4.1と4.2での例は、日中両言語は名詞に名詞、動詞に動詞ときちんと対応しているが、そうでない場合も少なくは無い。これは特に慣用表現の翻訳において多いと考えられる。

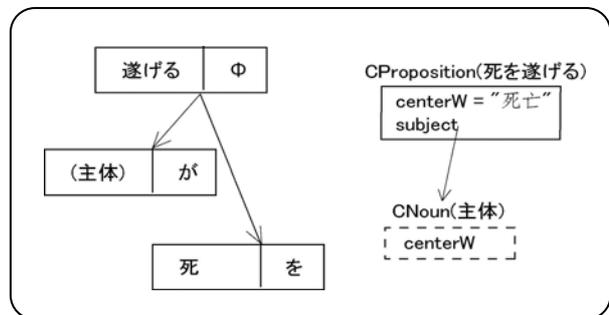


図6 「<主体>は死を遂げる」

例えば、「<主体>は死を遂げる」のような慣用句は中国語の「<主語>死亡」に訳す。図6は「<主体>は死を遂げる」の日中両言語の表現構造を示すものである。「死を遂げる」は中国語ではあわせて一単語の「死亡」に翻訳され、日本語の三節の構造から中国語の二節の構造になり、構造上にズレが生じる。jaw/Editorを用いて、このようなズレのある翻訳規則を書くことができる。

しかし、「彼は悲惨な死を遂げた」という文の場合は、中国語ETに「死」に対応するオブジェクトがな

いため、日本語側の「悲惨だ」のような「死」に修飾する要素は中国語側では係り先が無くなり、文章全体のETの変換に失敗する。このような場合に対して、「死」を修飾する要素のため、ET側で「死」に対応する空のオブジェクトを係り先として確保しておけばよい[3]。但し、[3]ではこの空オブジェクトに係る形容詞を副詞に翻訳するだけの場合を解決したが、他の連体修飾は解決していない。例えば、「それが彼の心を引く」のように、「心」に係る名詞「彼」があれば、中国語側では述語の「吸引(心を引く)」の直接目的語に変えて翻訳する場合もある。

文字面節のどの修飾要素をどのように変えるかという情報は日本語表現に依存し、予想できるので、予めその日本語表現の翻訳規則に書き込むのは可能である。例えば、「死」に係る形容詞は、中国語側では副詞に変えるので、空オブジェクトに

「adjectival2="adverb"」という変換情報を持たす。「心」に係る名詞は中国語側では直接目的語に変えるので、空オブジェクトに「m_of2="direcObject"」という変換情報を持たす。ズレのある表現ごとに適当な情報を与えると、ズレを分類する必要はなくなる。

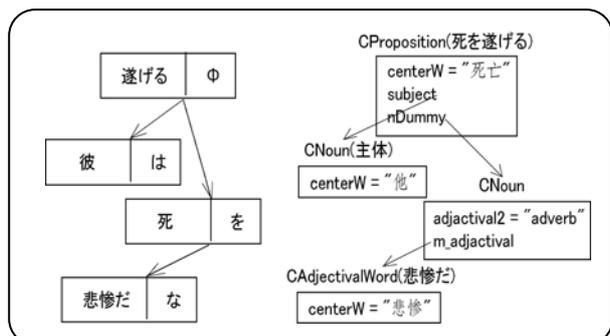


図7 「彼は悲惨な死を遂げた」修正前のET

現在のjaw/Chineseでは、空のオブジェクトに記述した情報によりそれに係る要素を中国語らしく変えETを修正し、空のオブジェクトを解放し、修正したETを線状化する手順で構造上にズレが生じた日本語表現を翻訳する。

図7は「彼は悲惨な死を遂げた」全体のETへ変換

できたものを示す。「死」に対応する名詞の空のオブジェクトはnDummyメンバが指し、「死」に係る形容詞は空のオブジェクトのm_adjectivalメンバが指す。図8は「adjectival2="adverb"」という変換情報により修正した「彼は悲惨な死を遂げた」の全体のETを示すものである。

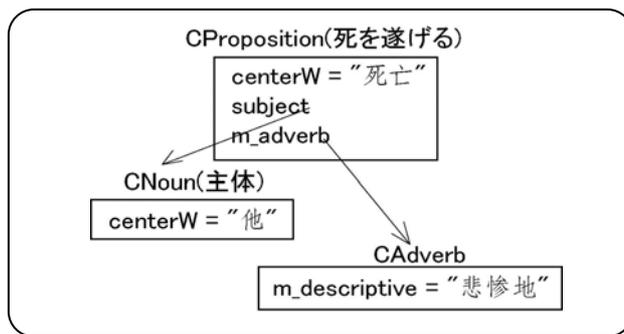


図8 「彼は悲惨な死を遂げた」修正後のET

5 おわりに

日中機械翻訳システムjaw/Chineseによる接続表現の処理について述べた。接続表現のある日本語の翻訳には概ねに対応できる。但し、接続が2箇所以上になる場合、「あの借家人は行って取り立てないと家賃を払ってくれません。」のように、「ない」の否定範囲を判断するなどの問題が起こってくる。今後の課題である。

また、自立語に文字面条件を用いることでjaw/Chineseでは翻訳の精度をある程度高めたと考えるが、構造上にズレのある日中両言語表現の翻訳に関しては、さらに数多くの翻訳実験で検証していく予定である。

参考文献

- [1]日本語英語の文対応データベースの作成, 村上仁一、池原悟他, 第7回 LACE 研究会, pp1-10, 2002.12.1.
- [2]日本語から多言語への翻訳エンジン jaw, 宇野修一、福本真哉、田中伸明、松本忠博、池田尚志, 言語処理学会第11回年次大会, pp538-541, 2005.
- [3]機械翻訳システム jaw と多言語への翻訳実験, 浅井良信、宇野修一、卜朝暉、福本真哉、池田尚志, 言語処理学会第11回年次大会, pp237-240, 2005.