

対応する表現が構造的にずれている場合のjawによる翻訳処理について

—jaw/Chineseにおける事例—

穆貴彬 宇野修一 福本真哉 浅井良信 卜朝暉 池田尚志
岐阜大学 工学部

1. はじめに

翻訳では、原言語と目的言語の間で、名詞が名詞に、動詞が動詞に、副詞が副詞にといったように同種の構造に対応させられない場合、あるいはそのような対応では不自然で無理な対応になるといった場合がたくさんあり、機械翻訳における課題の一つである。我々は日本語を原言語としていろいろの言語に翻訳する機械翻訳システム jaw を開発しているが[1,2]、本報告では、そのような構造にずれのある表現に関するいくつかの事例（中国語への翻訳での事例）を取り上げ、jaw/Chinese による対処法について述べる。

2. jaw の概要

jaw 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相手言語の表現構造を対応させるパターン翻訳知識ベースと、相手言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[1,2]。このほか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[3]。相手言語の表現構造は、3節で例示するような各種の表現要素がリンクで結ばれたネットワーク構造であり、C++のオブジェクトの構造として実装している。パターン翻訳規則は、日本語の表現パターンに対応する相手言語の表現構造を作り出すプログラムとしてデータベース化されている。このように相手言語の表現構造（オブジェクト）を介して翻訳を行う点が jaw の特徴であり、クラスを適切に設計することが重要である。

構造のずれに関する対処も、適切なクラスメンバを設定し必要な情報を書き込むこと、表現構造を適切に作成すること、適切な線状化関数を構築すること、という中で行われる。

3. 日中言語間での構造のずれへの対処の事例

3.1 「夢にも V (ない)」

これは「夢にも」という名詞文節が、中国語では

「夢を作っても」という意味合いの動詞文になる例である。

ここで君に会おうとは夢にも思わなかった。

即使 做梦 也没 想过 在这里 遇见你。

でも 作る 夢 でもない 思う た ここで 会う 君。

「夢にもV (ない)」という表現はjaw/Chineseでは、「夢」をキーワードとするAddition_CW型のパターンとしてリレーショナルデータベースのテーブルに記述されている。Addition_CW型は内容語をキーワードとし、「その内容語を含む文節がどのような文節に係るか」を記述するためのパターンである。表1に簡略化したテーブルの一部を示す。

表1 「夢にも V ない」の日本語パターン

Type	BID	KID	KW	FW	SA	FW
Add_cw	1	2	夢	にも		
	2	0			[用言]	ない

表1でTypeはパターンのタイプ、BIDは文節番号、KIDは係り先の文節番号、KWは検索の対象とするキーワード、FWは機能語に関する条件、SAはその文節の内容語に対する（意味的な）制約条件を表している。

この日本語パターンに対する翻訳規則で作成される中国語の構造を表現するオブジェクトの構造（表現構造）を図1に示す。

図1には、4つのブロックがあるが、それぞれはC++のオブジェクトを意味する。「CProposition

(做)」のオブジェクトは述語のクラスCPropositionのオブジェクトであり、centerW、directObjectはそのメンバの一部であり、値はそれぞれ“做”という訳語、「CNoun (夢)」のオブジェクトへのポインターであることを表している。「CpConnection (即使…也…)」のオブジェクトは、“夢を作る”という従属節を“Vない”という主節に接続するための表現に関わる要素をメンバとする接続詞のオブジェクト

である。connectP1とconnectP2は、接続に対応する訳語“即使”(connectW1)と“也”(connectW2)を配置する位置を指示するメンバである。CProposition (V)のブロックは、主節の動詞(「夢にも」に係る動詞)を表している。CProposition (V)のメンバCModeJは、日本語の用言後接機能語に関わる翻訳処理のために設けてあるメンバであり、主節 (V) の訳語の生成の過程で、別に記述されている「ない」に関する翻訳処理がなされることになる。

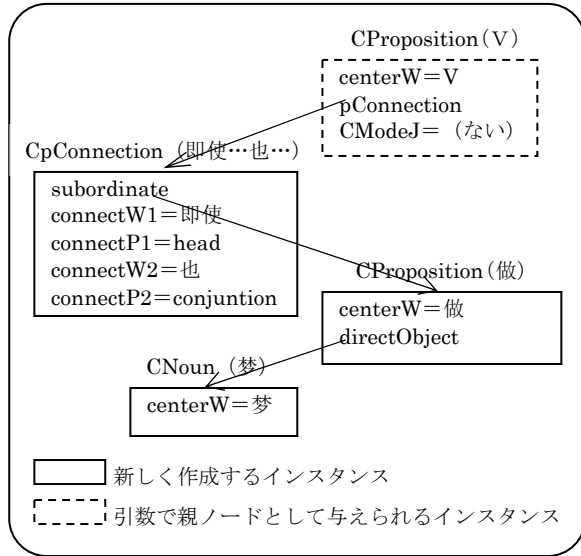


図1 「夢にも V (ない)」に対応する中国語の表現構造

表現構造の各クラスは目的言語の1次元言語表現を生成するための線状化関数(生成関数)をクラスのメソッドとして持つ。線状化関数をルートノードから再帰的に呼び出すことで訳文が得られる。「夢にもVない」のパターンは線状化関数によって「即使 做梦也 V (ない)」のように線状化される。

3.2 「Nのあおりを食ってV」

これは「Nのあおりを食って」という動詞文が、中国語では「Nの原因で」という意味合いの名詞文になる例である。

交通ストのあおりを食ってこのあたりはどの道路も混雑している。

因为 交通罢工 这 附近 无论 哪条 道路 都 拥挤。

原因 交通スト この あたり どの 道路 も 混雑している。

「Nのあおりを食ってV」という表現は

jaw/Chinese では、「食う」をキーワードとする Addition_CW 型パターンとして記述されている。

「あおり」は字面照合される。表2に簡略化したテーブルの一部を示す。

表2 「Nのあおりを食ってV」の日本語パターン

Type	SA	TXT	FW	KW	FW	SA
	[体言]		の			
		あおり	を			
Add_cw				食う	て	
						[用言]

TXT は字面照合される部分を記入するフィールドである。

この日本語パターンに対する翻訳規則で作成される中国語の表現構造を図2に示す。この表現構造は「因为 N V」のように線状化される。

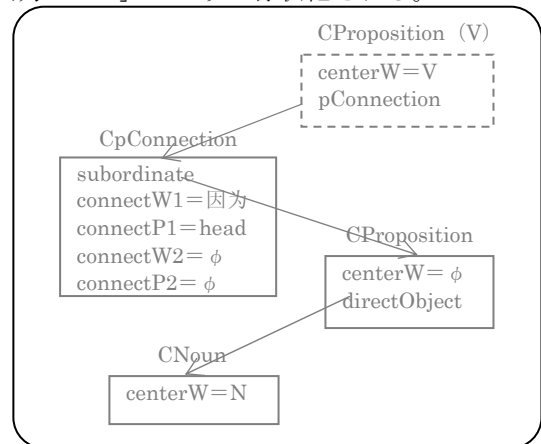


図2 「Nのあおりを食ってV」に対応する中国語の表現構造1

この表現に対応する中国の表現は「因为N」という名詞フレーズである。したがって、図2のようなCpConnectionを用いた接続型の表現構造ではなく、図3のように普通に「因为N V」という名詞フレーズがVを直接修飾する構造として表現の方が素直であるし、それで正しく翻訳することができる。しかし、入力日本語文が「Nのあおりをまともに食ってV」のように「食う」を修飾する別の要素が現れた場合には、図3の構造では対処できない。そこで、このような表現に対しては図2のように日本語の構造と相似した構造を用いて対処することとしている。図2のCPropositionは訳語(centerW)を持たない特殊なオブジェクトである。適切に設計した線状化関数によって「Nのあおりをまともに食って

V)のように「食う」は「正因为N V」という正しい中国語の表現を得ることができる。

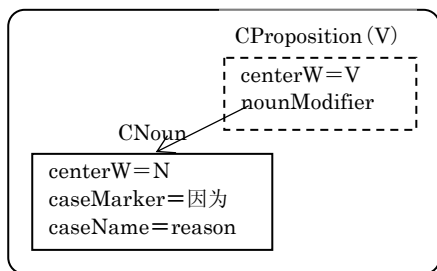


図3 「Nのあおりを食ってV」に対応する中国語の表現構造2

3.4 「N2を泡立てる」

日本語の単独の動詞で表現される意味が中国語では動詞と補語によるフレーズとして表現される場合がしばしばある。この例はその一つである。

卵白を泡立てるにはふつう泡立て器を使う。

要 打 蛋清 打 得 起泡 通常
使用 打蛋器。

には 混ぜる 卵白 混ぜる 泡立つ ふつう
使う 泡だて器。

「N1がN2を泡立てる」の日本語パターンを表3に示す。

表3 「N1がN2を泡立てる」の日本語パターン

Type	BID	KID	SA	FW	KW
	1	3	[人]	が	
	2	3	[食料]	を	
Base	3	0			泡立てる

日本語の動詞「泡立てる」は中国語では動詞「打」（混ぜる）とその補語「起泡」（泡立つ）という2つの表現要素を使った表現となる。中国語の表現構造を図4に示す。補語に対応する CComplement というクラスを設計した。

この表現構造は線状化関数によって中国語の補語表現に使う助詞「得」が自動的に加えられ、中国語の程度補語の文法に沿って「打」を2回繰り返し、「N1打N2打得起泡」のように線状化される。

「N1がN2を激しく泡立てる」のように、「泡立てる」を修飾する表現が加わった場合でも、それは「CProposition (打)」が用意しているメンバに入って問題なく翻訳できる。「N1がN2を激しく泡立てる」は「N1 剧烈地 打N2 打得 起泡」のように線

状化される。

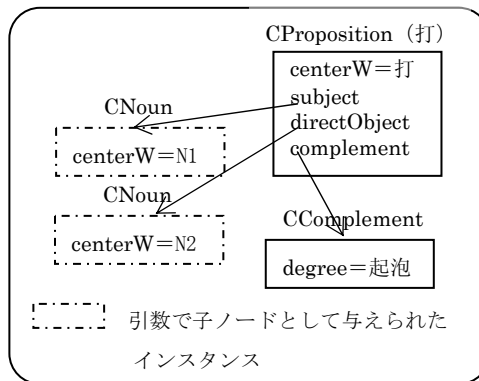


図4 「N1がN2を泡立てる」に対応する中国語の表現構造

3.5 他の翻訳ソフトとの比較

前述した事例を市販の翻訳ソフトで翻訳してみた。

- **ここで君に会おうとは夢にも思わなかった。**

市販の翻訳ソフトでは「在这里打算遇到你一点也没想」と訳される。「夢にも」は日本語の副詞の一語として扱い、中国語の副詞「一点也」（少しも）と翻訳しており、意味は通じるが、日本語の「夢にもVない」の意味が素直に訳されていない。この他、「会おう」が「打算遇到」と訳され、意外の「Vよう」が計画の「Vようと思う」と訳されてしまっている。文全体の語順も違っており、「ここで君に会おう」は「思う」の目的語として「想」の後ろに来るべきである。

- **交通ストのあおりを食ってこのあたりはどの道路も混雑している。**

市販の翻訳ソフトでは「吃交通罢工的冲击这个附近哪个道路也混乱着」と訳される。「Nのあおりを食ってV」は直訳になっており、意味が通じない。この他、指示詞がよく訳されていない。指示詞の訳語が1文字（「この」は「这」で、「どの」は「哪」である）の場合、指示詞と名詞の間にその名詞に対する数量詞（「あたり」に対する数量詞はないが、「道路」に対する数量詞は「条」である）が必要で、「指示詞 数量詞 名詞」という構造になるべきだ。取立て詞の「も」もよく翻訳されていない。

- **卵白を泡立てるにはふつう泡立て器を使う。**

市販の翻訳ソフトで「要弄起沫卵白一般弄起沫使用容器。」と訳される。「泡立てる」を「弄起沫」の

一語に訳しており、動詞の「弄」と補語の「起沫」の間に入るべき目的語が「弄起沫」の後ろに来てしまっている。その他「卵白」と「泡だて器」の訳語の間違いもある。

以上のように、市販の翻訳システムは言語間の表現に構造的なずれがある場合に適切な翻訳処理がなされていないように見受けられる。目的言語の表現構造を介して翻訳処理する **jaw** の手法はこれまでの事例では有効であった。さらに翻訳例を重ねて検証していきたい。

4. jawEditor について

jaw はまず目標言語の表現構造を作成し、次いで表現構造に記述された情報をもとに目標言語の1次元の文を生成するが、前節のいくつかの事例で示したように両言語の表現のずれは目標言語の表現構造を経由すること、適切な線状化関数を設計することで吸収することができる。

このように **jaw** では翻訳規則として表現構造を自在に設計できることが重要となるが、それを支援するシステムが **jawEditor** である。

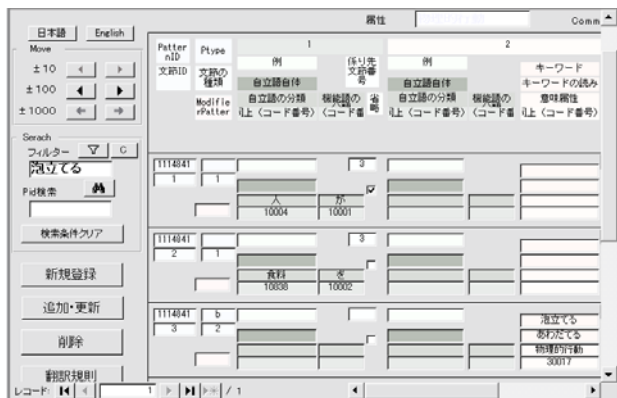


図5 「泡立てる」の日本語パターンの記述

jawEditor は日本語パターンを記述・編集する画面(図5)と、それに対応する翻訳規則を記述・編集する画面(図6)がある。翻訳規則の作成画面では規則のタイプとクラスのメンバの設計にそって、前節で述べたような任意のオブジェクトのネットワーク(現在のところ、リンクの深さは5階層までとしている)に対応する翻訳規則を作成することができる。

翻訳規則は翻訳規則関数への変換モジュールによってC++の関数に変換される。作成された翻訳規則関数はコンパイルされ、それぞれの目的言語専用の

DLL となって管理される。翻訳規則関数を実行すると目的言語の表現構造がオブジェクトのネットワークとして作り上げられる。**jawEditor** は、この翻訳規則関数のDLLを作成することまでを行う。

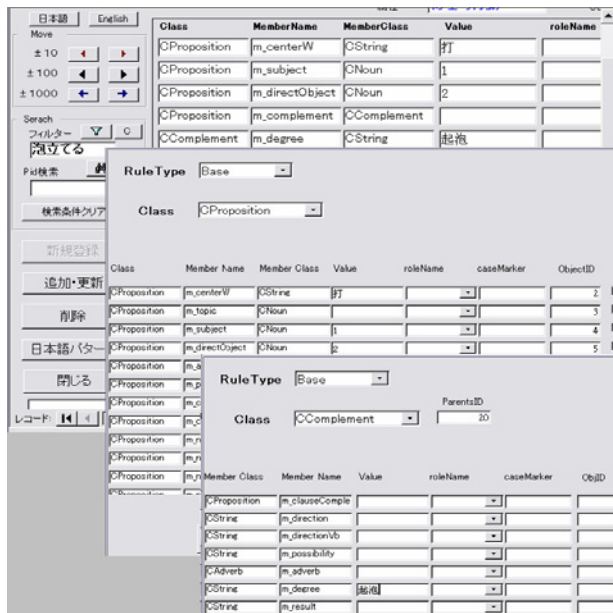


図6 「泡立てる」の翻訳規則の記述

5. おわりに

原言語と目標言語間で表現の構造にずれがある場合でも、目標言語の表現構造のクラスとそのクラス・メソッドである線状化関数を適切に設計することで、問題をうまく吸収できることを、**jaw/Chinese** におけるいくつかの事例で示した。今後はさらにさまざまな対訳例文を収集し、システムおよび規則の整備・強化を続けていく予定である。

参考文献

- [1] 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, 他, 言語処理学会第11回年次大会, pp233-236, 2005
- [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] 日中機械翻訳システム **jaw/Chinese** における取立て表現の翻訳処理, 田中伸明, 卜朝暉, 浅井良信, 宇野真哉, 伊佐佐和哉, 松本忠博, 池田尚志, 言語処理学会第11回年次大会, pp189-192, 2005
- [4] 日中機械翻訳システム **jaw/Chinese** における変換・生成の方式, 謝軍, 今井啓允, 池田尚志, 自然言語処理Vol.11, No.1 p.43-80, 2004