

日中機械翻訳システム jaw/Chinese における 取り立て表現の翻訳処理

田中申明, 卜朝暉, 浅井良信, 宇野修一, 伊佐治和哉, 松本忠博, 池田尚志
岐阜大学工学部

1 はじめに

我々は, 日本語からいろいろの言語へ翻訳するパターン変換型機械翻訳エンジン jaw の開発を行っており, 現在は中国語, シンハラ語, ベトナム語, ミャンマー語および手話への機械翻訳を試みている.

本研究では, 日本語における取り立て表現の翻訳処理を jaw に実装した. 具体的には, 先に言語学的立場から分析していた日中言語間での取り立て表現に関する翻訳規則 [1] を jaw に組み込み, jaw/Chinese で取り立て表現の翻訳が行えるようにした.

2 機械翻訳システム jaw

jaw は VC++ で記述されたパターン変換方式の機械翻訳システムである. 「入力文の木構造 (Input Tree・IT) の作成」「翻訳規則の木構造 (Transfer Tree・TT) への変換」「目的言語の表現構造木 (Expression Tree・ET) の生成」「ET の各オブジェクトの線状化 (目的言語文の生成)」の 4 つの処理で目的言語への翻訳を行う. 図 1 に jaw の基本的な処理の流れを示す.

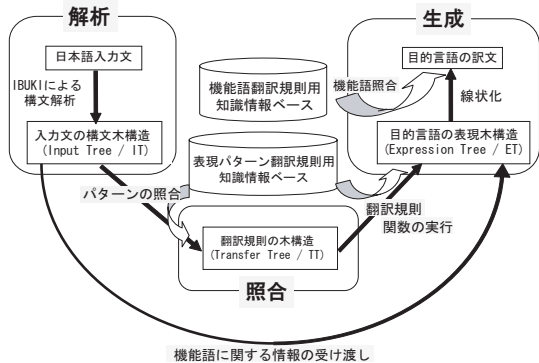


図 1: jaw の処理の流れ

3 日中言語間における取り立て表現の対応関係

3.1 取り立て詞について

取り立て詞とは「文中のさまざまな要素を取り立て, これとこれに対する他者との関係を示す」作用を持つ助詞である. 以下に例を示す.

(1). 先生 も 今日のパーティーに出席した.

(2). 戦争中, 16 歳の子供 も 徴兵された.

(3). ご飯 さえ 食べられれば満足だ.

(4). 誰 も 私の話を聞いてくれない. (1) は, 「先生」

以外の人もパーティーに出席したことの暗示, (2) は, 意外の意味, (3) は, 最低限の十分条件, (4) は, 全面的な否定, を表している. そして「も」は, (1),(2),(4) のように複数の意味用法を持つ.

3.2 取り立て表現の対応関係

取り立て表現の翻訳規則については [1] で分析した. 表 1 に「さえ, も, でも」についての意味分類の一部を, 表 2 に訳語選択規則の一部を, 表 3 に語順規則の一部を示す.

表 3: 取り立て詞の訳語の語順処理規則 (一部)

訳語 T'	語順処理規則
連…都	『 <u>連</u> 』: X' が主語であればその直前. その他ならば「 <u>連</u> + X'」を主語の直後. 『 <u>都</u> 』: 状語/謂語の直前.
甚至	状語/謂語の直前.
既…又	『 <u>既</u> 』: X1' の直前, 『 <u>又</u> 』: X2' の直前
也	主語があればその直後に, なければ状語/謂語の直前に置く.
都	X' が賓語以外ならば, X' の直後. X' が賓語ならば, 「X' + 都」を謂語の直前.
即使…也	『 <u>即使</u> 』: X' の直前. 『 <u>也</u> 』: 状語/謂語の直前.

4 取り立て表現の翻訳処理

4.1 jaw での機能語の翻訳処理

jaw では機能語に関する翻訳規則は, 日本語の表現パターン (命題的内容) に対する翻訳規則の実行によって作成される ET (目的言語の表現構造木 (C++ のオブジェクト)) に対して, 必要な情報を追記する形の規則となっている.

規則の記述は, 機能語に対してそれを目的言語で表現するために必要となる言語表現上の情報をフィールドとするテーブル上に記述する. テーブルは一対一で一意的に翻訳できる場合と一対多で訳し分けが必要となる場合に分けている. 一対多の場合は, ET から必要となるメンバーの値を参照することで訳し分けを行うが, その条件記述もテーブル上でやっている.

表 1: 取り立て詞の意味分類 (一部)

取り立て詞	意味分類	中訳語	例文
さえ	意外の意味	連…都	自分の名前さえ書けない。
		甚至	太郎は授業中に寝てさえいた。
も	列挙	既…又	彼女は親切でもあり、裕福でもある。
	他者に対する肯定	也	私も東京へ行く。
でも	全面的な肯定	都	どこでも行ける。
	X でも結果は変わらないという譲歩の意味	即使…也	先生でもこの問題は解けない。

表 2: 取り立て詞の訳語選択規則 (一部)

取り立て詞	日本語パターン	X の属性	P の属性	X の中訳語の属性	訳語
さえ	X さえ P		仮定形以外	主語/賓語/状語/補語	連…都
				謂語	甚至
も	X1 でもあり X2 でもある	形容動詞		謂語	既…又
		名詞			也
でも	X でも P	疑問詞	肯定		都
		名詞	確言/完了		即使…也

上記のように記述された規則を用いた翻訳処理の流れを以下に示す。

1. 一対対応の訳語選択規則テーブルを参照し、訳語を決定する。
2. 複数の訳語候補が存在する場合は、一対多対応のテーブルを参照し、様々な条件の中から ET の情報にマッチする条件を見つけ、訳語を決定する。

用言後接機能語に関しては、この処理は既に実装され翻訳実験も行ってきたが、取り立て詞 (体言その他に後接する) に関しては未実装であった。

4.2 取り立て表現の翻訳処理の実装

上述した用言の場合と同様の方法で表 2, 3 の規則を組み込み、取り立て詞に対する翻訳処理を実装した。

ここで、取り立て詞『も』を含む「私も東京へ行きたい」という文に対する翻訳処理の例を図 2 に示す。ここでは取り立て詞『も』と用言後接機能語『たい』に対する翻訳処理が行われている。以下で詳しい処理の流れを説明する。

(1) 一対対応の訳語選択規則から用言後接機能語『たい』に関する規則を参照する。『たい』に対する中訳語は一意に決まるため、助動詞を格納するメンバ変数 (m_aux_verb) へ訳語「想」をセットする。

(2) (1) の用言の場合と同様に、取り立て詞『も』についても一対対応の訳語選択規則を参照しに行くが、『も』は複数の訳語候補を持つ取り立て詞であるため、ここでは訳語を決定することができない。この

ような場合は、次に参照すべき一対多対応の訳語選択規則テーブル名 (図では『も』に関する規則テーブル名"SP_mo") が記述されているため、次にそのテーブルを参照する。

(3) 訳し分けに必要な様々な条件について ET の情報と照合を行う。照合はテーブルの先頭から行い、同じ"ID"を持つ全てのレコードの条件を満たしたとき、そこに記述された訳語を選択する。この処理について以下で更に詳しく説明する。(ここで、『も』に取り立てられている要素を T と表すことにする。この例では T=「私」となる。)

- ID=13 の条件についての照合

Sub_ID=1 では、T が主語にあたるかを判定。結果は真であるため、次に Sub_ID=2 で T の品詞カテゴリが数量 (m_category=60) かを判定。T の品詞は名詞 (m_category=1) であるため結果は偽となり、ID=13 の条件については成り立たない。

- ID=14 の条件についての照合

Sub_ID=1 で T が主語にあたるかを、Sub_ID=2 で T の品詞カテゴリが名詞かを判定。どちらも真であるため、ID=14 の条件を満たし、ここに記述された訳語「也」が取り立て詞『も』の訳語として選択される。この場合「也」は語順規則により、主語の直後に置く、となっているため主語の直後に訳語を線状化するメンバ変数 (m_postSubj) へ「也」をセットする。

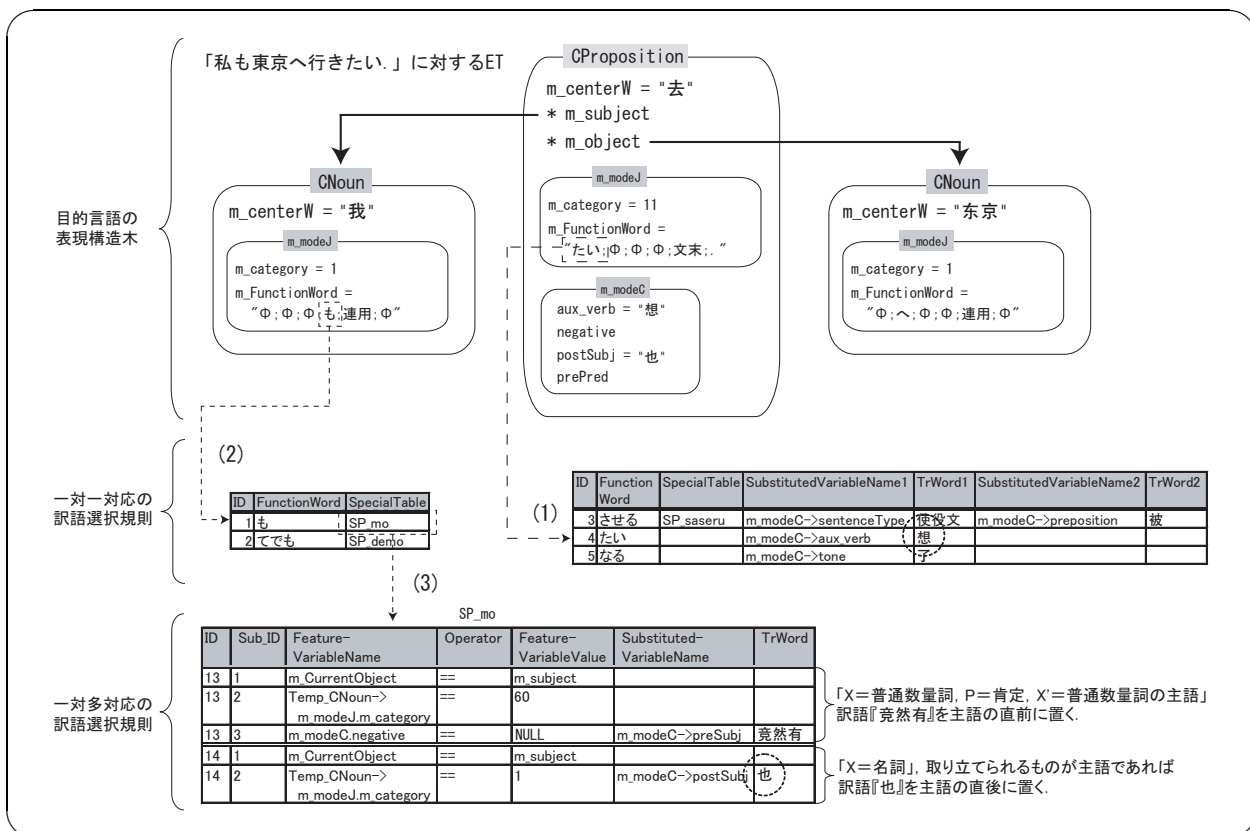


図 2: 機能語に対する翻訳処理の例

4.3 翻訳実験と考察

4.2 節で実装した翻訳処理が、取り立て詞の翻訳に対して有効であるかどうかを検証するため、取り立て詞の意味分類(表 1)に用いた例文(14 文)を対象として翻訳実験を行った。その翻訳結果の一部を表 4 に示す。また、この翻訳結果に対して、(評価 1): 訳文として正解かどうか(全文比較)、(評価 2): 取り立て詞の翻訳に着目し、その訳語・語順が合っているかどうか、の 2 点について著者のうちの中国語母語話者が評価した。

評価 1 (jaw)

全文比較に対する評価の内訳を表 5 に示す。

表 5: 評価 1 (jaw)

翻訳結果の内訳	該当数
< 正解 >	13
x < 一部誤り >	1

『x』となった文(ID=4)は「家から駅まで...」という場所の始点・終点に関する語順が誤っていた。これについては、今後語順について検討を進めれば解決できると考えられる。

また、ID=1, 7...などの機能語「られる」を含む文については、可能・受身などの曖昧さの処理について jaw ではまだ完全に実装しているわけではないが、ここではデフォルトの訳「能」を出力して正解と評価している。

評価 2 (jaw)

取り立て詞の翻訳に対する評価の内訳を表 6 に示す。

表 6: 評価 2 (jaw)

翻訳結果の内訳	該当数
< 正解 >	14
x < 不正解 >	0

14 文全ての場合で取り立て詞に対する訳語・語順が正解となっており、取り立て表現に対する翻訳処理はうまく機能したと考える。

次に市販の日中機械翻訳ソフトでも同じ 14 例文に対して翻訳を行い、上記と同様に評価 1, 2 を行った。この翻訳結果を表 4 に、評価の内訳を表 7, 8 に示す。

表 4: 例文に対する翻訳結果 (一部)

ID	例文	翻訳結果 (上段: jaw, 下段: 市販のソフト)	評価 1	評価 2
1	自分の名前さえ書けない。	<<连>>自己的名字<<都>><<不>><<能>>写		
		连自己的名字不能写	x	x
2	私は金も暇もない。	我钱<<和>>时间<<都>>没有		
		我钱也空暇也没有	x	x
3	五人も遅刻した。	<<竟然有>>五个人迟到<了1>		
		五人迟到	x	x
4	家から駅まで三時間もかかる。	从家<<竟然>>花三小时到车站	x	
		从家到车站三小时花费	x	x
5	千円もあれば足りる。	如果有一千日元就够了		
		一千日元有够	x	
6	何人も遅刻した。	<<好>>几个人迟到<了1>		
		好几个人迟到		
7	誰もが信じられる。	谁<<都>><<能>>相信		
		谁都能相信		
8	誰も信じられない。	谁<<都>><<不>><<能>>相信		
		谁都不能相信		
9	私も東京へ行く。	我<<也>>去东京		
		我也去东京		
10	どこでも行ける。	哪里<<都>><<能>>去		
		哪里都能行	x	
11	どんな人でもできない。	什么样的人<<都>><<不>>能		
		无论怎样的人不能	x	x
12	先生でもこの問題を解けない。	<<即使>>老师<<也>><<不>><<能>>解答这个的问题		
		老师也不解开这个问题	x	x

評価 1 (市販ソフト)

表 7: 評価 1 (市販ソフト)

翻訳結果の内訳	該当数
< 正解 >	5
x < 一部誤り >	9

『～た』のような「時制」を表す訳語がなかったり、適切な訳語が選択されていないものがあつた。また、中国語の語順としては本来 SVO の形をとるべきであるが、ID=4 の文については SOV の形になっており、中国語の訳文としては不適切である。

ここで、jaw との比較の意味で機能語「られる」を含む文に着目すると、全て同じ訳語「能」が出力されていた。

評価 2 (市販ソフト)

表 8: 評価 2 (市販ソフト)

翻訳結果の内訳	該当数
< 正解 >	8
x < 不正解 >	6

取り立て詞の訳語が出ていないものも何文があつたが、訳語が出ていても適切なものが選択されていないものがあつた。例えば、「意外性の意味」を表す訳語とすべき文が「他者への肯定」を表す訳語をとって

しまつたり、「提案か例示」が「他者への肯定」になっていたりした。これでは、本来の文の意味を損なつてしまい、取り立て詞の翻訳に関しては不適切である。

5 おわりに

本稿では、取り立て表現に関する翻訳処理を jaw へ実装したことについて述べた。実装実験には、中国語を目的言語とした jaw/Chinese を用い、日中機械翻訳における日本語の取り立て表現と中訳語との対応を規則化したデータを組み込んだ。例文に対して翻訳を行ったところ、取り立て詞に対する翻訳は正しく行われており、取り立て表現の翻訳方式についての手がかりを得ることができた。また、市販のソフトとの比較も行ったが、市販のソフトでは取り立て詞の翻訳に関しては誤りがあつた。

今後は、更に多くの例文に対して翻訳を行い、取り立て詞に対する規則や翻訳処理の追加を行う予定である。

参考文献

[1] ト朝暉:『日中機械翻訳における取り立て表現の翻訳について』自然言語処理 Vol.9 No.5, Oct, 2002
 [2] 今井 啓允:『日本語からアジア諸言語への機械翻訳の試み』情報処理学会第 65 回全国大会講演論文集, March, 2003