

日中両言語の構文上の対応にずれがある表現の翻訳に関する分析 — 日中機械翻訳の角度から —

† 朝暉 † 浅井信良 † 王軼謳 † 穆貴彬 † 池田尚志 †

† 岐阜大学工学部・学振外国人特別研究員/中国広西大学外国語学部 † 岐阜大学工学部

E-mail bu@ikd.info.gifu-u.ac.jp

1. はじめに

翻訳を行う際、原言語と目的言語の対応に構文上でずれがある表現が存在している(連体構造を連用し、名詞を動詞に翻訳する等)。このようなずれ表現の翻訳は人間翻訳にも機械翻訳にも難点である。言語学の側では人間による翻訳に、増訳、減訳、転(変)訳、分訳などの手法を用いて対応するというような論述があるが[1][2][3]、機械翻訳にとっては考慮しなくてもよい内容も混在しているため、機械翻訳の観点からの整理と洗練が必要になる。本論文では機械翻訳の角度から、日中両言語の構文上の対応にずれがある表現の中の、受動態/能動態のずれ、品詞のずれ、および訳語の数のずれに関わる表現を中心に考察し、日中機械翻訳のための翻訳方法を提案する。またわずかなサイズではあるが、まとめた受動態/能動態のずれ分類に属する文の出現率を調査し、本規則による翻訳精度を手作業で評価した。

2. 構文上の対応にずれがある表現とその分類

2.1 ずれがある表現の定義

本稿で議論する「構文上の対応にずれがある表現」は日中両言語では通常それぞれ対応する構文上のジャンルがあるが、場合によって通常のジャンルで対応できず、異なるジャンルの表現で翻訳するものである。例えばある条件の下で、受動態を主動態に、連体修飾構造を連用修飾に、名詞を動詞に翻訳するなど。議論の対象は機械翻訳であり、構造を変換してもしなくても翻訳できる場合は、容易な方(変換しない翻訳)を採用する。また文学翻訳の「更なる自然な文、美しい文に翻訳する」というような目標は求めない。以下この方針で議論を進めていく。

2.2 ずれがある表現の大別分類

文献[1][2][3]などで述べている人間翻訳の日中翻訳手法(増訳、転訳、減訳など)を参考にし、機械翻訳の必要も考慮して、我々は日中両言語の構文上の対応にずれがある表現を下記のように大別する。

[A] 相(態)のずれ [変訳]

(1) 自然言語処理学会は来月に開かれる。

→ 自然语言处理学会下个月召开。

(日: 受動態 ⇄ 中: 能動態)

[B] 品詞のずれ [変訳]

(2) 大勢の人が困っているから何か起こったに決まっている。

→ 因为有很多人围住, 一定是发生什么事了。

(日: 動詞 ⇄ 中: 副詞)

[C] 成分のずれ [変訳]

(3) この国は石油資源が豊富にある。

→ 这个国家有丰富的石油。

(日: 連用修飾語 ⇄ 中: 連体修飾語)

[D] 訳語の数のずれ

[D-1] 単語の数の増加 [増訳]

(4) 私はおじいさんから面白い話を聞いた。

→ 我从爷爷那里听说了有趣的事。

(場所名詞「那里(ところ)」の増加)

[D-2] 単語の数の減少 [減訳]

(5) すぐ来るから、待ってください。

→ 我马上来, 请等一下。(「から」の訳語の減少)

[E] 修飾成分の語順のずれ [変訳]

(6) 高校時代の友達の多くはすでに結婚している。

→ 小学时的许多朋友, 已经结婚了。

(日: 友たちの多く ⇄ 中: 多くの友たち)

[F] 単文と複文のずれ [変訳]

(7) 私は彼を私の恩人と考えている。

→ 我认为他是我的恩人。

(日: 目的語は名詞句 ⇄ 中: 目的語は小文)

3. ずれ表現の分析と翻訳

2節で分類した「転訳」手法で対応する[A]「態のずれ」と[B]「品詞のずれ」、「増訳」手法で対応する[D-1]「訳語の数の増加」の表現を中心に分析する。「態のずれ」は能動態と受動態のずれを考察する。以下の分析は文献[1][2][3]などを参考にし、また著者の翻訳の経験からの分析に基づいて、機械翻訳の角度からまとめたものである。

3.1 受動態と能動態のずれに関する分析と翻訳

日本語の受動態は基本的に「N1はN2にVられる/れる」のように表現されるが、中国語では動詞の変形がなく、「N1被N2V」という表現が対応する。「られる/れる」は中国語では基本的に受動態のマークの前置詞の「被」で対応できる。しかし日中の能動態と受動態の使用のすべてが一致しているわけではない。そのずれは次の三種類に分けられる。全体では中国語の受動態の使用は日本語ほど頻繁ではないと見受ける([i]は[iii]より多い)。

[i] 日: {受動態} ⇒ 中: {能動態} (られる = Φ)

[ii] 日: {受動態} ⇒ 中: {特殊受動態} (られる = 「由/受」等)

[iii] 日: {能動態} ⇒ 中: {受動態} (Φ = 被)

3.1.1 日: {受動態} ⇒ 中: {能動態} / {特殊受動態} の場合

日中の能動態と受動態におけるずれの[i]と[ii]の場合の翻訳規則を表1のようにまとめた(表1のパターンでは「られる」のみ表示するが「れる」のことも代表する)。

3.1.2 日: {能動態} ⇒ 中: {受動態} の場合

日本語の能動態を中国語では受動態に翻訳する場合は少ないが、日本語の動詞が能動態で受動の意味を表す場合、中国語では受動態に翻訳することがある。即ち、N1はN2にV→N1被N2V。例えば、

(8) 顔は炎に赤く染まっていた。→ 脸被火光染红了。
 (9) 彼女はこの映画に感動した。→ 她被这个电影感动了。
 分析：このような場合、N2は無生物で、Vは自動詞である。他に「石につまづく、雪に埋もれる」などがある。
 このような日本語動詞は受動態になれず、動詞自身が受

動の意味を含んでいる。機械翻訳では個々の動詞のパターン規則の記述で解決できる。

規則例： 「つまづく」
 <N1>は<N2>につまづく < N1=人/動物 N2=具体物>
 → N1 被 N2 绊倒

表1 日本語の受動態を中国語の主動態及び特殊の受動態に翻訳する規則(「られる」に対応する中訳を中心に)

規則番号	翻訳条件		「られる」の中訳	例文
	日本語パターン構成	構成要素の条件		
(1)	NはがVられる	a N=抽象的な物事 b V=抽象性の動作を表す他動詞	Φ	①自然言語処理学会は来月に開かれる。 →自然语言处理学会下个月召开。
(2)	N1には/にN2がVられている	a N1=場所(<地名/組織名/場所/場>) b V=他動詞	Φ	②彼の涙にはある種の期待が込められている。 →他的眼泪包含着某种期待。
(3)	(a) (Nは/が) PとVられる (b) (Nは/が) PものとVられる	V=「思う,考える,見る,見なす,推測する,言う,伝える,見込む,予想する,見積もる,話す,語る,推定する」などの伝達動詞。P=節	Φ	③中国は人民元を更に値下げするだろうと伝えられている。 →据说中国可能再贬值人民币。 ④会談は成果が挙げられないものと見られる。 →看来会谈不会取得成果。
(4)	N1に/でVられた/られているN2	a. N1=範囲/場所 b. V=他動詞	Φ	⑤壁に掛けられている絵はとても美しい。 →挂在墙壁上的画很美。
(5)	NはVられた/ているものだ		Φ	⑥この木は去年植えられたものだ。 →这棵树是去年种的。
(6)	(a) N1は(Adv)Vられてある (b) N1はN2にVられてある	V=他動詞 N2=場所	Φ	⑦家具はきちんと並べられてある。 →家具摆的整整齐齐。 ⑧離婚届はテーブルの上に置かれてあった。→离婚申请放在桌子上。
(7)	(a) N1が(は)N2にVられる (b) N1はN2にVられる	N1=人/動物 (a) V=「死ぬ/泣く/倒れる/来る/なる」などの自動詞(「降る/吹く」は除外) N2=省略不可 (b) V=他動詞	Φ Φ	⑨花子がおかあさんに死なれた。 →花子死了妈妈。 ⑩ おかあさんが子供に泣かれた(困っている)。 → 孩子哭, 妈妈很难办。
(8)	N1はN2に言われてN3をV	V=他動詞	让	⑪兄に言われてこれをやった。 →哥哥让我做这个事。
(9)	N1はN2に見舞われる		遭受	⑫今年陽子江の南部は大きな水害に見舞われた。 →今年长江南部遭受了大水灾。
(10)	「N1はN2に付き添われる/守られる」		由	⑬彼女は両親に付き添われ病院へ検査に行った。 →她由双亲陪同去医院做检查了。 ⑭大統領の専用は戦闘機三機に守られて空港に着陸した。 →总统的专用机由三架专机守护着降落在机场。

説明:1. (1)-(7)は[i]に、(8)-(10)は[ii]に属するずれである。(8)-(10)は動詞ごとのパターン翻訳規則で訳すことになる。
 2. (1)には、抽象名詞はNTT語彙大系を参考するが、抽象性の動詞の属性の特定は困難な場合がある。
 3. (2)と(4)の規則のN1の属性として、抽象的な場所/範囲の限定は機械には困難な場合がある。
 4. (7)番は間接受動文に関する問題である。(a)には動詞ごとにパターン規則で翻訳できるが(b)の動詞の数は推測し難く、また訳語も特定しにくいと、課題である。

3.2 品詞のずれに関する分析と翻訳

A. 名詞を動詞に翻訳する(コプラ文に関連するずれ)

(A-1) 動作性を表す名詞を動詞に翻訳する

(10) 私は来月アメリカへ行く計画だ。

→ 我计划下个月去美国。

分析：(10)のようなコプラ文は中国語では通用しない。「計画」に対応する名詞訳語もあるが、この場合は「…と計画する」のように、動詞として翻訳すべきである。パターン翻訳では「だ」をキーワードとし、「だ」の名

詞部を字面条件とする次のパターン規則で解決できる。

<N1>は<V><計画>だ (N1=人/動物) (V=行動)

→ N1 计划 V

名詞「計画」は中国語の述語動詞として訳し、「だ」に対応する訳はない(だ=Φ)。「アメリカへ行く」の部分は連体修飾語から動詞フレーズの目的語に変換する。

また、「計画」の連体修飾語は動詞ではなく、「名詞+の」のような構造に変わっても相似する中国語訳になる。

<N1>は<N2>の<計画>だ (N1=人/動物 N2=動作性名詞)

→ N1 计划 N2

この場合のN2はほとんどサ変名詞でもであると推測する。
また、「計画」以外に、「つもり/予定」などの動作性名詞も上記と同じの構造では同様な翻訳になる。

(A-2) 思想, 気持ちを表す名詞を動詞に翻訳する。

(11) そのとき苦しい思いだった。

→ 那个时候觉得很苦。

分析: 「苦しい思いだった」は「苦しく感じた」に翻訳する。次のパターン翻訳規則で解決できる。

(Nは) <adj> < > 思い > だ → N 觉得 <adj>

B. 形容詞(近い, 強い)を動詞に翻訳する

(12) 大会は終わりに近い。→ 大会接近结束。

(13) 北極熊は寒さに強い。→ 北極熊耐寒。

分析: (12)の形容詞の「近い」は中国語では動詞「接近(接近する)」に、(13)の「強い」も中国語では動詞「耐(耐える)」に翻訳する。翻訳規則として形容詞ごとに「終わり」の部分の字面条件とするパターンを登録すればよい。

<N>は<終わり>に近い → N 接近结束

C. 動詞を副詞に翻訳する

(14) 大勢の人が困んでいるから、何か起こったに決まっている。

→ 因为有很多人困住, 一定是发生什么事了。

(15) 太郎はいつも一つ事を繰り返して話す。

→ 太郎经常重复说话。

(16) 何十回の登りを経て、とうとう山頂に辿り着くことに成功した。

→ 经过几十回的攀登, 终于成功地到了山顶。

(17) 私は一貫して彼を支持する。

→ 我一贯支持他。

(18) 悪天候の中太郎は苦労して車を運転した。

→ 在恶劣的气候中, 太郎辛苦地开着车。

分析: (14)は動詞「決まる」を副詞の「一定(必ず)」に翻訳する。翻訳規則としては次のパターンで翻訳できる。

<用言>に決まっている → 一定 <用言>

(15)は動詞の接続形「繰り返して」を中国語の副詞の「重複」に翻訳する。次のパターン規則で翻訳できる。

<N>を繰り返して <V2> → 重複 V2N

(16)-(18)はサ変動詞の問題でもある。(16)は文末で現れ、

(17)-(18)は「して」接続形として現れる。翻訳規則としてはそれぞれのサ変動詞をキーワードとするパターンの構成で翻訳できるが、数が多いと予測され、また同じサ変動詞でも「運動して汗をかく」のような場合は、副詞ではなく動詞に翻訳したりすることもあるため、サ変動詞の属性から訳しわけの条件に関する考察が必要である。

D. 動詞を形容詞に翻訳する

(19) 正常では空気中の酸素と二酸化炭素の配分率は決まっている。

→ 正常情况下, 空气中的氧和二氧化碳的分配比率是一定的。

分析: 「決まる」は日本語では動詞であるが、中国語では「一定的」という形容詞に翻訳し、コプラ動詞「是」も用いる必要がある。動詞をキーワードとするパターンの構成で解決できると思う。

パターン翻訳には、上記の品詞のずれに関しては、基本的に各構成要素の属性を工夫した用言のパターン規則

で解決できるが、このようなずれがある動詞/形容詞などがどのようなものであるかについての分析は課題である。

3.3 訳語の数のずれの増訳に関する分析と翻訳

ここでは日本語の人称省略に対する増訳は対象としない。

A. 述語を共有する並列文の増訳

(20) この大学では成績はA, B, C, Dの4ランクに、卒論は優秀、合格と不合格の3ランクに分けている。

→ 在这个大学, 成绩分为A, B, C, D四个等级; 毕业论文分为优秀, 合格和不及格三个等级。

分析: 日本語文の述語は文末にあり、例(20)のような並列文では、異なる補語は一つの述語「分ける」を共有することができるが、中国語ではそれぞれ二つの述語「分爲」として訳す必要がある。機械翻訳では中国語文を生成する段階でもう一つの述語の補完処理が必要である。

B. 「Nの立場(方面/角度/視点)から」の連用構造の増訳

(21) 機械翻訳の立場から、このような文学的な美しい翻訳を追求するのは未だ困難である。

→ 从机器翻译的角度(来)看, 要追求这样优美的文学性的翻译还是困难的。

(22) 被害者の立場からこの問題を考えよう。

→ 从被害者的立场来考虑这个问题吧。

分析: 「…立場から」という構造は中国語では状語(連用修飾語)になり、それを受ける述語「看, 考虑, 考察(見る/考える/考察する)」などのいわゆる伝達動詞の増訳が必要になる。ただし、例(22)のように、述語動詞がすでに伝達動詞であれば、中国語文では増訳する必要がない。機械翻訳では(21)のような場合は次のような規則で翻訳できる。(例(22)の場合は普通のパターンで翻訳する)。

<N>の立場/角度/視点から P → 从 N 的立场来看 P

P = 「見る/考える」等の伝達動詞以外の述語

C. 「N1からのN2」の連体修飾構造の増訳

(23) シベリアからの風は我慢できないほど冷たい。

→ (23-1) 从西伯利亚吹来的风冷得难以忍受。

→ (23-2) 来自西伯利的风冷得难以忍受。

(24) そのとき、彼からの手紙は私の励ましであった。

→ 那个时候, 他的信给了我勇气。

(25) 小学生時代からの夢は一瞬で消えてしまった。

→ 从小学生时代就拥有的梦想一瞬间消失了。

分析: 「N1からのN2」の構造は例(23)と(25)のように、中国語では「N1からVN2」のように動詞Vを増加して翻訳する場合がある。但し(24)のように「から」を削減して翻訳する場合もある。また(23)の場合には、(23-2)は「N1から来たN2」のように、固定的に動詞「来」+前置詞「自」を使用する。(23-1)の訳では、「N1から吹いて来たN2」のように訳し、それも正解であるが具体的な動詞「吹く」の増加が必要になり、N2からどのような動詞の増加が必要かを判断することとなるため、パターン規則を複雑化し、パターンの数も増える。従って、機械翻訳には(23-2)の訳で対応したほうがよいと考える。

(23)には次の(a)のようなパターン規則で翻訳できる:

(a) <N1>からの<N2> N1=場所 N2=体言

→ 来自 N1 的 N2

(24)には次の(b)パターン規則で翻訳する:

(b) <N1>からの<N2> N1=人/動物 N2=体言

→N1 的 N2 (「から」を訳さないようにする)
(25)には(c)のように N2「夢」ごとにパターンを構成する(「N1 からの」を字面条件とする)ことで解決できるが、パターンの数が多いと予測する。

(c) <小学生時代からの>夢 → 从小就拥有梦想

D. 場所詞、方位詞の増訳

(26) ゴムのりはテーブルにおいてある。

→ 胶水放在桌子上。

(27) 母から面白い話を聞いた。

→ (我) 从妈妈那里 听说了有趣的事情。

分析: (26)の場合には中国語では「テーブルの上におく」というように、方位詞「上」を増加して翻訳する。(27)では「母から」を「母のところから」というように場所詞「那里」を増加した。このような場合は動詞パターンの翻訳規則から方位詞/場所詞の補完処理が必要である。文献[6]では我々のシステム jaw/Chinese における場所詞、方位詞の増訳処理を述べたことがある。

E. 「N1 による N2」の連体構造の増訳

(28) 去年日本では大雪による災害がひどかった。

→ 去年在日本, 因为大雪引起的灾害很严重

(29) エジソンによる発明は数多くある。

→ 爱迪生的发明有很多。

分析: この構造は中国語では「N1 によって VN2」のように動詞を増加して訳す場合がある。(28)のような N1 が N2 の「原因」を表す場合大抵「引起/产生」の動詞を増加するが、「原因」の属性の付与は機械には困難である。(29)では N1 が N2 の動作主の場合、「N1 の N2」のように翻訳し、動詞の増訳がない。

F. 日本語で使用されている語と語の組み合わせが中国語では適応できない場合の増訳

(30) 掲示板に合格者が発表された。

→ 在布告板上公布了合格者的名单。

分析: 中国語では「合格者」と「発表」は直接に動目構造の組み合わせとしては使用できない。「合格者のリストが発表された」というように「名单」の訳語を増加して動詞「公布」と組み合わせる。このような組み合わせの動詞と名詞は数的には多くはないと予測し、下記の字面条件を含む動詞のパターン規則で翻訳できると考える。

<N1>に<合格者>が発表される N1=場所

→ 在 N1 公布合格者的名单

G. 「かなりの N」の構造の増訳

(31) 彼は不動産業でかなりのお金を儲けた。

→ 他在不动产业上挣了相当多的钱。

分析: この構造は中国語では「かなり多くの/大きな…N」のように訳し、量を表す形容詞の増加が必要である。次のようなパターン規則で翻訳できるが、具体的にどの形容詞を補充すればよいかは N の属性から推定できるかと思うが、それに関する詳細な考察は課題である。

<かなり>の<N> (N=体言) → 相当多/大…的 N

H. 「相当に V」の構造の増訳

(32) 彼は相当に泣いた。→他哭得相当厉害。

(33) 中国語の本は相当にある。→有相当多的中文。

分析: このような構造は中国語では「相当にひどく泣く」、「相当に多くある」というように、「相当」を修飾する形容詞「厉害/多」などを増加して翻訳する。また(32)では

「相当厉害」の部分は中国語の補語にり、(33)のような述語が存在動詞の場合、「相当多」が連体修飾語になる。動詞ごとのパターン(「相当に」を字面条件とする)の構成で解決できるが、中国語生成のときそれぞれ補語あるいは連体修飾語として生成する処理が必要である。

4. 評価

わずかなサイズではあるが下記の二つの評価を行った。

(1) 本論で列挙したずれがある 46 の日本語例文をある市販翻訳ソフトで翻訳したところ、わずかな 30%の正訳率しか得られなかった(ずれ部分のみの評価)。

(2) 表 1 で示しているずれがある受動文の出現率と表 1 で提案した翻訳規則に基づく翻訳精度を手作業で評価した(表 2)。同時に、市販翻訳ソフトの結果とも比較した(表 3)。(両方とも「られる」に対応する訳語の正確性のみを注目して行った)。評価のデータはある日英対訳コーパス[5]の中から抽出したものである。

表 2 表 1 のずれ受動文の出現率とその規則による正訳率

対象文	抽出した受動文	ずれがある受動文の出現率	ずれがある受動文の正訳率
10000	760	211/760=28%	159/211=75%

表 3 市販ソフトのずれ受動文の翻訳結果

ずれ受動文	正訳文	正訳率
211	102	48%

表 2 の評価では、現在のずれ分類と現規則では対処できない例などがあつた。主な問題点は文献[4]で述べた。

5. 終わりに

本論文は日中機械翻訳のために、両言語の構文上の対応にずれがある表現の中の、受動態と能動態のずれ、品詞のずれ、訳語の数のずれの 3 種類を中心に分析し、その翻訳方法を提案した。またわずかなサイズではあるが、まとめた受動態と能動態のずれについて手作業によってそのずれ文の出現率と本翻訳規則の精度の評価を行った。760 受動文に対して約 28%の出現率と約 75%の正訳を得て、市販の低い正訳率と比較して、本研究の必要性と有効性を一定の程度で確認した。本論で分析したずれの分類とその内容は未だ不完全なものであるが、このようなずれ表現のデータ収集と分析は日中機械翻訳の問題発見と翻訳レベルの向上に大きな意義があると考えられる。

謝辞: 本研究は日本学術振興会外国人特別研究員事業の支援により実施したものである。

参考文献:

- [1] 刘绩生编著, 实用日汉对译技巧, 湖南人民出版社, 1997.
- [2] 苏琦编著, 汉日翻译教程, 商务印书馆, 1998.
- [3] 今富正巳, 新訂中国語⇔日本語翻訳の要領, 光生館, 1973.
- [4] 卜朝暉他, 日中機械翻訳における構文上の対応のずれに関する考察—受動態と能動態のずれ、品詞のずれを中心に—, 情報処理学会研究報告 2006・NL-176, pp33-40, 2006.3.
- [5] 村上仁一, 池原悟他, 日本語英語の文対応データベースの作成, 第 7 回 LACE 研究会, pp1-10, 2002, 12.1.
- [6] 宇野修一, 卜朝暉, 今井啓允, 池田尚志, 日中機械翻訳システム jaw/Chinese における連体修飾表現、場所表現等の処理言語処理大会第 10 回年次大会発表論文, pp640-643, 2004.3.