

日英両語における論述様態の対応関係—言語情報文法の観点から

On the correspondence in the linguistic utterances between Japanese and English—from a viewpoint of the linguistic information syntax

岩垣 守彦
Morihiro Iwagaki

フリー
Freelance

All linguistic information can be considered to be equally composed of a single verb and a noun/ nouns, which can be called 'unit-information' and in case two or three pieces of information are needed, they are connected with connectives, such as conjunctions, prepositions, and verbal. So, if unit-information and connectives are changed properly, linguistic information of one language can be changed into another one.

I would like to make a proposal concerning an idea of converting Japanese into English, unit-information for unit-information.

1. はじめに

ある小説の中に、北海道の列車に乗って外を眺めている場面があって、

やがて森が切れ、海が見えてきた。

という文が出てくる。日常よく使われる表現であるが、市販の翻訳ソフト、web 上の機械翻訳にかけてみると、正しく変換できない。1947 年ごろに英仏の単語の置き換えで始まり、日本でも 1955 年以降続いているコンピュータによる言語変換が今でもできないとなると、かつて MIT のパパート博士が言ったように、機械翻訳は時間と金の無駄なのであろうか。翻訳家が日本語から英語へ翻訳する際に使う手法、つまり、従来の単文・重文・複文という区分ではなく、「単位情報と連結の対応」という手法から再検討する必要があるように思われる。

そもそも、言語情報というものは「名詞」と「動詞」を組み合わせた「単位情報(日本語: (whn)·(hw)·(whr) + N + V, 英語: N + V + (whn)(whr)(hw))」が情報の基本単位であり、必要に応じてそれを二つ以上「連結」して発信されるものである。言語の個別性は「符牒」(signs), 「単位情報内の符牒配列」(unit-information patterns), 「語りの型」(narrative patterns), 「連結の仕方」(connection patterns)がそれぞれ異なることから生じる。したがって、これらの対応関係を適正に処理すれば、言語の相互変換は可能と思われる、しかし、実際に変換が出来ないのは、個々の言語に関する研究は進んでも、「言語情報の対応」という観点からの研究が進んでいないからと思われる。ここでは、言語

情報の対応関係を「単位情報(N + V + whn·whr·hw)」と「連結」という観点から見て、「言語情報の変換」に関わる要素を検討してみる。

2. 言語変換に関わる要素

上に示した日本語符牒の文は

時(やがて)
+ 単位情報(森が切れる)
+ 連結(,)
+ 単位情報(海が見えてくる)

と、単位情報と連結の組み合わせからできている。

まず、「日本語符牒を英語符牒に変えて」、「単位情報内の符牒配列を整える」と

時(before long)
+ 単位情報(there be break in forest)
+ 連結(,)
+ 単位情報(sea come into view)

となる。それから、「日本語の連結(,)を英語の連結(and)に対応させ、「語りの型」と「現実時間との関わり(今以前のこと)」の要素を加味すると

時(before long)
+ 単位情報(there was a break in the forest)
+ 連結(and)
+ 単位情報(the sea came into view)

となる。まとめると、

やがて森が切れ、海が見えてきた。

↓

Before long there was a break in the forest, and the sea came into view.

と変換することができる。つまり、一つの言語情報を他の言語情報にできるだけ等価的に変換するには、基本的には、

「符牒(signs)」を対応させる。

「単位情報(unit-information)」を対応させる。

「連結の型」(connection patterns)を対応させる

「語りの型」(narrative patterns)に対応させる

「現実の時間との関わり」を対応させる。

という五つの処理を適正に行なえなくてはならないことになる。

3. 変換に必要な要素の対応

「符牒」と「単位情報」の対応

「符牒(signs)」の対応に関しては、機械翻訳のはじめから考えられていて、池原悟(2005)によると、「日英機械翻訳システム ALT-J/E(池原ほか 1987)において、単文構造を対象とした表現意味辞典「日本語語彙体系」(池原ほか 1997)が開発されている。この辞書は、動詞と格要素との関係を 1,7000 件の結合値にまとめたものである。格要素となる名詞の意味的な用法は、名詞意味属性(2700 分類)を用いて指定されており、40 万語の名詞意味辞書と連動する。翻訳実験では、IPAL (IPA 1987) の例文 5000 件に対して、90%の正解率が得られており、方式限界は 97%であることが報告されている。(金出地ほか 2001)これにより、日英機械翻訳における単文の意味的な訳し分けの問題はほぼ解決できたとみられるので…(pp.1-2)とのことである。

また、情報に必須な要素の「時(when)、場(when)、方法・様態(how)」などを組み合わせた「単位情報(unit-information)」の対応に関しては、切れ目なく符牒が並ぶ日本文の構造分析の道具として「形態素解析」が考案されており、符牒の関係性を捉える道具として「係り受け解析」がある。これらは複雑な日本文の構造を確定させるのに有用であるように見受けられる。たとえば、

A やがて森が切れ、海が見えてきた。

B やがて森が切れて、海が見えてきた。

を、形態素解析は「切れ」を「基本連用形」、「、」を「特殊1 読

点2」とし、「切れて」を「タ系連用テ形」、「、」を「特殊1 読点2」と解析する。さらに「係り受け解析」(京大の juman|knp)にかけると、

A やがて———

森が——— |

切れ、<P>

海が——— |

見えて来た。<P>-PARA

B やがて———

森が——— |

切れて、——

海が—— |

見えてきた。

と区別する。しかし、言語情報としては、どちらも同じで、

時(やがて) + 単位情報(森が切れる)

+ 連結(、て、)

+ 単位情報(海が見えてくる)

↓

時(before long) + 単位情報(there be break in forest)

+ 連結(and)

+ 単位情報(sea come into view)

である。

「連結」の対応

言語情報を「単位情報」+「連結」+「単位情報」…ととらえると、等価的に変換するには「連結の型」(connection patterns)を確定して対応させなければならない。この点に関しても、たとえば、池原悟の「非線形性に着目した言語表現モデルと重文と複文に対するパターン辞書の開発」(第10回 LACE 研究会(2005.12.24-25)に基づいて重文・複文を集めた『鳥バンク(Tori-Bank)』という労作がある。しかし、これには「単位情報」と「連結」という考えは導入されていないし、英語の連結の重要な要素である「情報の比重」も考慮されていない。

あらためて「連結」という観点から日英両語の連結関係を考え見ると、単位情報の連結は「情報の比重」に基づく連結、つまり、

日本語 [単位情報] + [連結辞・接続詞] + [単位情報]

英語1 [未了解単位情報] + [接続詞 + 未了解単位情報]
 英語2 [接続詞 + 了解情報] + [未了解情報]

のような対応関係も含めて、7種類に分類することができる。

0「単位情報」

付: 単位情報内の「符牒連結」

0日1「名詞」[連結辞(の, か, と, など)]「名詞」(連結辞)

0英1 (prep. +) N + and, or, of, with, by, in, etc.) + N

0日2「形容詞・副詞」[連結辞(て, が, など)]「形容詞・副詞」

0英2 adj./ adv. + and/ but + adj./ adv.

0日3「動詞」[連結辞(して, て, など)]「動詞」

0英3 verb + by/ in + ing-form

(注:「…(し)て(から)…」は2単位情報文である)

0「単位情報の並置」(「単位情報」→「単位情報」→「単位情報」…)

1「単位情報」の内的連結

内的連結とは、一つの単位情報の名詞・代名詞が他の単位情報と動詞を介して共有関係をもつことで、

A [動作主体と共有関係]

「主体 + …(し)に、…するために(は)、…(する・なる)まで、…して(いて)、…すると + 動詞、など」

B [名詞と共有関係]

「(いつも)…する…している…した…する + 名詞」
 (主に連体節)

の二つの関わり方がある。これらの要素を加味して「連結の型」を分類すると、次の通りである。

「単位情報」の内的連結1

[通時]

1B1「…する…である + 名詞」

Noun + (...ing) (=which + present tense)

1B2「…られ(て)いる + 名詞」

Noun + p.p. (=which is p.p.)

[同時]

2A1「…し(ながら)…すると…したまま + 主語(名詞・代名詞)」

…ing……, mS (noun/pronoun)(passive voice は (being) p.p.)

2A2「～を…し(ながら)・したまま + 主語(名詞・代名詞)」
 mS + mV, (with) + noun + (...ing-form/ p.p.-form)

2B「…する・している + 名詞」

Noun + ...ing (=which is + present participle)

[以前]

3A「…した(ことがある) + 主語(固有名詞・代名詞)」

Having + p.p./ ...ing + mS(noun/pronoun)

3B1「…(され)た + (無意志)名詞」(受動態)

Noun + p.p.-form

3B2「…した + 名詞」(能動態)

Noun + (relatives) + past-tense

[以後]

4A1「…する(ため)に(は)…(求め・欲し)て…(し)に」

mS + mV + to root-form

4A2「…する(ため)には…」

(In order) to root-form, mS + mV

4B1「…する(ことになる)…する(という) + 名詞」

Noun + to root-form (=which is to root-form)

4B2「…する(ための) + 名詞」

mS + mv + noun + to root-form

「単位情報」の内的連結2(主観連結)

5日1「…と言う、思う、信じる、わかる、など」

5英1 Subjective view (say/think: etc. + that; what; etc.) + info-unit

(参考: mS + subjective view + mV + subjective view + mS + mV)

5日2「…のようだ…で[で]うれしい・かなしい・遺憾に思う、など」

5英2 It seems to-root form

S + V + adj. + to-root form

5日3「(行ってみると)Oは…だ(った), Oが…る, Oを…させる・してもらう、など」

5英3 S + V + O + root form, ing-form, to-root form, p.p.-form, etc.

「単位情報」の外的連結1(未了解情報 + 接続詞)

6日 [単位情報] + [連結辞・接続詞] + [単位情報]

6英1 [未了解単位情報] + [接続詞 + 未了解単位情報]

6英1-1 [info-unit 1] + [and[while, when, as, etc.] + info-unit 2]

- 6英1-2 [info-unit 1] + [but + info-unit 2]
- 6英1-3 [info-unit 1] + [or + info-unit 2]
- 6英1-4 [info-unit 1] + [so + info-unit 2]
- 6英1-5 [info-unit 1] + [because[for] + info-unit 2]
- 6英1-6 [info-unit 1] + [then + info-unit 2]

「単位情報」の外的連結2 (接続詞 + 了解情報)
 7英2 [接続詞(When, If, Because, Since, etc. + 了解情報) + [未
 了解情報]

4. 「連結の型」(connection patterns)の運用例

[日本文]

犬は餌が欲しくて人に媚びてくるので、捨てられたばかりだとわ
 かった。

↓

「単位情報(主体の補充)の型」の対応

単位情報[(犬は:)餌が欲しい]

+ 連結[て]

単位情報[(犬は:)人に媚びてくる]

+ 連結[ので]

単位情報[(犬は:)捨てられる]

+ 連結[(だ)と]

単位情報[わかる].

↓

「符牒」の対応と「連結の型」(connection patterns)の対応

単位情報[(dog:) seek food]

+ 連結[て]→内的連結 4A1

単位情報[(dog:) make up to people]

+ 連結[ので]→外的連結 6英1-4

単位情報[(dog:) lost owner]

+ 連結[(だ)と]→主観連結 5英1

単位情報[it: suggested]

↓

The dog made up to people to seek food, so it suggested that the
 dog had lost the owner

↓

「修飾処理」(ばかり→only just)

↓

The dog made up to people to seek food, so it suggested that the
 dog had only just lost the owner.

↓

「代替処理」(文脈情報)(同一・関連語関係→代名詞・代動
 詞)

↓ ↓ ↓

its which it

↓

「変換文」 The dog made up to seek food, which suggested that
 it had only just lost its owner.

参考文献

池原悟「非線形性に着目した言語表現モデルと重文と複文に
 対するパターン辞書の開発」(第10回 LACE 研究会
 (2005.12.24-25)

岩垣守彦(1993-a)『英語の言語感覚---ルイちゃんの英文法』
 (玉川大学出版部, 1993)

岩垣守彦(1993-b)『日本人に共通する和文英訳のミス』(ジャパ
 ンタイムズ, 1993)

岩垣守彦(2003a)「機械翻訳の精度を上げるための構文解析の
 提案」(2003/03/25 自然言語処理研究会, 東京工科大学)

岩垣守彦(2003b)「日本文の「つなぎ」と英文の「つなぎ」の対応
 に関して」(2003/05/26 自然言語処理研究会, 東京工業大学)

岩垣守彦(2003d)「連体修飾節の英訳に関して」(2003/07/25 自
 然言語処理研究会, 山形大学)

岩垣守彦(2007)「日本語の論述様態の定型化について」(人工
 知能学会 workshop, 2007/07/21)

岩垣守彦「単位情報の観点から文型を考える」(電子情報通信
 学会・思考と言語研究会, 機械振興会館, 110204)

岩垣守彦「単位情報の内部構造」(1)―単位情報の観点から
 2011/08/26 ことば工学研究会, 神奈川大)

岩垣守彦「日英の等価変換のための「単位情報の内的連結」の
 分類」(2011/11/25「ことば工学」, 拓殖大八王子キャンパス)

岩垣守彦「複数の「単位情報」の連結について―言語情報文法
 の観点から」(2011/11/26「思考と言語」研究会, 早稲田大)

岩垣守彦「複数の「単位情報」を連結する際に関係する要素と
 連結の分類リスト」(2011/12/17 第16回 LACE 年次研究会, 桜
 美林大四谷校舎)

佐良木晶・宮澤織枝・新田義彦(2007)「シテ形用言接続句の対
 訳データ構築と日英機械翻訳の訳質改善」(言語処理学会第1
 3回全国大会, 2007/03, 龍谷大学)