

# 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

## - 心理表現の日英対照を中心に -

佐良木昌 宮澤織枝 新田義彦  
日本大学 東京チャオ 日本大学

**あらまし** シテ形接続の用言接続句を含む日英対訳データを、大規模コーパスから抽出し、用言意味属性およびシテ形接続の意味分類を付与した。シテ形接続については、4区分・26細目の範疇的意味に分類した。本データによれば、シテ形用言接続句と英文型との相関には有意な傾向があることが判った。したがって適切な英文型を選択することが機械翻訳の英訳精度改善の鍵である。本データを機械翻訳の辞書に登録することによって、適切な英文型での英文出力が可能になるが、辞書登録のための技術的な工夫が要ることが判った。

### 1. はじめに

本論文の目的は、接続助詞「て」を介して複数の用言が繋がるシテ形接続形式を対訳コーパスから抽出して意味的に分類し、意味分類に対応して英語文型を整理することにある。加えて、本対訳データを日英機械翻訳に実装することで翻訳英文の精度向上を目指す。

シテ形接続において、先行シテ節に続く主節の用言が格要素を伴わず用言同士が直に繋がる接続の形態のみを採りあげ、本形態をシテ形用言接続句と呼ぶ。シテ節の動詞を従動詞、主節の動詞を主動詞と呼び、二つの動詞が直に繋がる接続の形態を動詞接続句と呼ぶ。さらに、シテ形動詞接続句が文末である述定の形態のみ採りあげ、この接続句が連体修飾として体言に係る装定の形態は採りあげない。また、形容詞を含む形態については、限定的に述べるにとどめる。

シテ形接続の担う関係の意味が様々であることから、意味分類が重要課題である。仁田による分類(付帯状態・時間的継起・起因的継起・並列を大分類とする)に基づいて[1]、意味分類体系を作成した。四つの大区分に加えて下位分類を設けたが、詳細は別項に譲る[2]。

なお、本稿で採りあげる対訳データは、戦略的基礎研究推進事業の研究領域「高度メディア社会の生活情報技術」・研究課題「セマンティックタイポロジーによる言語の等価変換と生成技術」プロジェクトが編纂した日英文型パターンと例文を改編・編集したものである(CRESTデータと略称する)。また、本稿では、心理状態や感情行為、すなわち心的状態に関わる表現を採りあげ、これらを広く一括して「心理表現」と呼ぶ。

### 2. シテ形用言接続句対訳データの概要

CRESTデータから、シテ形接続形式を含む対訳データ約3万1千件を抽出し、これら抽出データのうち文末述語をなすシテ形接続1570件について、名詞変数のみを残したパターンを作成した。そのうち、シテ形動詞接続句は約610件である。形容詞を含む接続句については、

データを作成中である。

文末述語をなすシテ形接続1570件については、二つの用言が表す事態と事態との間の意味関係について、シテ形意味分類体系(4大区分・26細目)に基づく意味分類名が付与されている。これにより、多義性がきわめて大きいシテ形接続を意味的に分類することが可能になった。本対訳データでは、上記の意味分類された個々の対訳データに対して、和文の従動詞と主動詞に意味属性を与え、さらに、それらに対応する英文の表現を抽出して、これに同じく意味属性を与えた。先述の研究プロジェクトが策定した用言意味属性大系の定めによる。

### 3. シテ形接続を含む心理表現の日英対照

#### 3.1. シテ形接続の意味的な論理構成

心理を表すシテ形用言接続句は、その意味的な論理構成が、( )主節が心理を表す、( )シテ節が心理を表す、( )主節・シテ節ともに心理を表す、の3つに分かれる。次に、シテ形接続の意味分類大系にある心理表現に関わる分類項目は、「A 付帯状態」の「心的状態(A-2)」、「C 起因的契機」の「心因(C-6)」であり、かつ心的状態・心因ともにその心理主体の存在を前提としていることから、「シテ節が心理を表す」を、

- 1) 主体の行為・状態に伴う「心理状態・感情行為」
- 2) 主体の行為・状態を引き起こした「心因」

に分ける[3]。以上の区分から、用言接続句の意味的な論理構成は、5つに細分・整理できる(表1)。本分類により、動詞接続句を述語とする和文と英文とを、以下、対照させる。

表1 心理を表すシテ節・主節の論理的・意味的な構成

	シテ節	主節
	心理・感情以外	主体の心理・感情
1	付帯的心理状態	主体の行為・状態
2	心因	主体の行為・状態
1	付帯的心理状態	主体の心理・感情
2	心因	主体の心理・感情

### 3.2. 動詞接続句を述語とする和文と英文との対照

#### 3.2.1. パターン (主節が心理を表す)

パターンに対応する典型的な英文型は、

- 1) 心的状態を表す過去分詞(叙述形容詞相当)の述語文
- 2) 感情行為の動詞能動文などである。

主節が心理状態・感情行為を表すとき、シテ節は、その誘因やきっかけ、あるいは状況を表す。ここで「誘因」とは、原因や「心因」( 2・ 2)ではなく状況的起因の範囲に含めるものとする。主節の表す「感情動作・心的状態」の誘因となる「見て・聞いて・読んで」などを表すシテ節には、英文においては、1)では、述語が採る前置詞句 (be bored/disturbed/surprised + prep.)、不定詞句 (be disturbed to hear of)、on+動名詞句 (be so upset on hearing) など、2)では、前置詞+動名詞句 (amuse oneself by reading)、when 節を採る複文 (mourn when someone hear)、さまざまな形態に対応する。

一方、シテ節は、主節の心理状態・感情行為の原因や理由も表すが、シテ節・主節が表す二つの事象を、因果関係として認識していることを前提としている。この場合には、理由節 because clause の複文の事例がある。

#### 3.2.2. 1パターン (シテ節が付帯的心理状態を表す)

シテ形接続が「付帯的心理状態」を表すとき、対応する英語文型は、典型的には、

- 1) 動詞 + 副詞類
- 2) 主節 + 分詞構文 (複文) である。

「付帯的心理状態」を表すシテ節には、英文においては、感情副詞 (wearily) や、in+感情名詞 (anger, indecision, joy, madness, sorrow, surprise, terror, wonder) / with + 表情名詞句 (a smile, a look of annoyance)、あるいは雰囲気を表す名詞句 (a sad air) の形態が対応する事例が顕著である。

また、英語文型の異型パターンには、

- 1) be 繫辞 + 名詞句 (友達は共鳴して聞いてくれた。My friend was a sympathetic listener.)
- 2) 結果構文 (先生は笑って承知した。The teacher smiled her consent.)

などがある。「結果構文」[4]では、和文の従動詞が表す付帯的な心理が、英文では、文中核たる動詞、つまり、主たる感情行為として表現されている。

本パターンについて、61件のサンプルを採取し、文種を調べた(表2)。and 等位接続が 3.3%、複文・分詞構文が 18.1%であり、単文は、78.7%である。つまり、和文では二つの動詞で表現されるが、英語では一つの動詞で表現される事例が有意に多い。

#### 3.2.3. 2パターン (シテ節が心因を表す)

本パターンは、心因が何等かの行動あるいは状態を引き起こすという因果的な表現といえる。シテ節が心因を表すとき、対応する英文は、典型的には、

- 1) 心的状態を表す過去分詞(叙述形容詞相当)の述語文
- 2) 動詞 + 副詞類
- 3) 無生物主語・他動詞の構文
- 4) so that の構文 (複文)

1)の過去分詞述語文は「結果構文」である場合が多く、心因が述語で表され、結果の行為・状態が前置詞句で表される (be frightened/provoked/shamed into...)。

2)の文型では、動詞 + 前置詞 + 感情名詞のという定型化したパターンを採ることが多い。1パターンとは異なり、前置詞 in は僅少で at, of, with などを探る事例が多い。この傾向はリーチおよびスヴァルトヴィックの見解に合致する[5]。

また、4)の文型は、強い感情を起因としてある行動を起こすという事柄の表現の定型として認められる (too to の構文も同様)。3)の心因を表す名詞・動名詞節を主語とする他動詞構文も認められる。何れも心因が事を引き起こす因果的表現であるが、理由の副詞節を伴う事例はない。

#### 3.2.4. 1パターン (シテ節が付帯的心理状態、主節が心理を表す)

このシテ形接続に対応する英文では、

- 1) 心的状態を表す過去分詞 (叙述形容詞相当) の述語文
- 2) 動詞句 + 副詞類

2)の文型に於いて動詞が採る前置詞は、with を採る事例が多い (with anger, with surprise, with a vengeance) [6]。また、自動詞 + 補語の文型の事例もある。

#### 3.2.5. 2パターン (シテ節が心因、主節が心理を表す)

シテ節・主節ともに心理表現である場合、その多くが、シテ節が心因を表し、心因により生じた心理状態や感情行為を主節が表すという傾向が認められる。

- 1) 心的状態を表す過去分詞 (叙述形容詞相当) の述語文
- 2) 動詞句 + 副詞類
- 3) and の等位接続 (重文)

1)の事例には「結果構文」の場合が多く、また慣用句 be frightened out of one's wits もあり、2)の例には、慣用句で一種の結果構文である soften into tears もある。「結果構文」では、2と同じく心因が述語で表され、結果の行為・状態が前置詞句で表される。4)の等位接続は僅少であり(表3)、and が心因とその心理的な結果状態との因果関係を表す(表5の例文 2)。because clause の事例はない。

#### 3.2.6. 形容詞を含む接続句 (略) [6]

#### 4. 実験

Web で提供されている日英機械翻訳を使って、CREST データと同じ和文を英訳した。三つの異なるサイトの英訳結果を採り、その文型の傾向を調べた。英訳が誤訳であるか否かに関わりなく出力英文を文型別に集計した(表4)。かつ、うち一つについては、同じ機械翻訳システムを実験機に導入して実験を行った。シテ形用言接続句に関する対訳データに基づいて動詞接続句をユーザ辞書に登録した後、英訳を実行し、辞書登録による英訳の改善について検証した。サンプル総数36、内訳は8, 1<sup>1</sup>, 5<sup>2</sup>, 5<sup>1</sup>, 3<sup>2</sup>である。

##### 4.1. 日英機械翻訳の英訳傾向

Web1 では、and 等位接続 / because 節を主な文型として。Web2 および Web3 では、多くが and 等位接続の文型であった。まれに、become + 補語(Web1)や、動詞 + 前置詞 + 感情名詞の文型を出力することもある(Web1: run away in surprise; Web3: escape in wonder)。無生物主語の事例は無い。Web1 では、理由節の複文の事例があった。他動詞構文、結果構文のような文型は出力されることは、見られなかった。

一方、比較した CREST データでは、and 等位接続は一例のみ、because clause によって心因を表す事例は僅かであり、著しい対照をなしている。

表4 機械翻訳の英訳文型の傾向

総数36 内訳は8, 1<sup>1</sup>, 5<sup>2</sup>, 5<sup>1</sup>, 3<sup>2</sup>

翻訳サイト	数	and 等位				動詞 + 副詞類			
		1	2	1	2	1	2	1	2
Web1	36	16	(2,10,2,1,1)	8	(0,3,2,1,2)				
Web2	36	31	(6,12,5,5,3)	0	(0,0,0,0,0)				
Web3	36	27	(6,11,4,4,2)	3	(0,2,1,1,0)				
CREST	36	1	(0,0,1,0,0)	22	(2,12,3,3,2)				

  

翻訳サイト	数	過去分詞述語 +				従属節			
		1	2	1	2	1	2	1	2
Web1	36	4	(2,0,0,2,0)	5	(4,0,1,0,0)				
Web2	36	2	(2,0,0,0,0)	0	(0,0,0,0,0)				
Web3	36	2	(2,0,0,0,0)	0	(0,0,0,0,0)				
CREST	36	6	(4,0,0,1,1)	4	(2,1,1,0,0)				

##### 4.2. 複合動詞相当として辞書に登録

動詞接続句については、動詞接続句全体を一語としてみて、複合動詞相当としてユーザ辞書に登録した。各パターンについて事例を、和文、辞書登録内容、初期の英訳、登録後の英訳、CREST データの模範訳の順にて示す(表5)。

表5 動詞接続句の辞書登録による訳質改善

和文	その子供は絵本を読んでおもしろがった。
Lexicon	*読んでおもしろがる/V/R5/ +1/HUM/が 2/XART/を =1 amuse/V/ "oneself by reading" 2
Before	I read, and the child was amused by a picture book.
After	The child amused oneself by reading a picture book.
CREST	The child amused itself by reading the picture book.
和文 <sub>1</sub>	父は共鳴して聞いてくれた。
Lexicon	*共鳴して聞いてくれる/V/UL/ +1 が =1 be/V/IT="was a sympathetic listener", "...
Before	Father was resonant and heard it.
After	Father was a sympathetic listener.
CREST	My father was a sympathetic listener.
和文 <sub>2</sub>	ピーターは驚いて硬直した。
Lexicon	*驚いて硬直する/V/S1/ +1/HUM/が =1 frighten/V/PAS/ stiff
Before	Peter became stiff in wonder.
After	Peter was frightened stiff.
CREST	Pete was frightened stiff.
和文 <sub>1</sub>	スーはほっとしてため息をついた
Lexicon	*ほっとしてため息をつく/V/K5/ +1/HUM/が =1 give/V/ "a sight of relief"
Before	Sioux was relieved and sighed.
After	Sioux gave a sight of relief.
CREST	Sue gave a sigh of relief.
和文 <sub>2</sub>	母は感動して涙した。
Lexicon	*感動して涙する/V/S1/ +1/HUM/が =1 soften/V/ "into tears"
Before	Mother was impressed and shed tears.
After	Mother softened into tears.
CREST	My mother softened into tears.

#### 5. 実験結果についての論議

表5に示した事例では、一般論として言うにはサンプル数が不十分であるが、<単純な文型・簡潔な英文>に改善されていること、これを着目点として指摘できる。英文の訳質は模範訳と同じ文型であり、構造的に改善されている。and 等位接続の文型は採らず、単文を中心としている。

辞書登録上の課題については、

- 1) 過去分詞述語 + 前置詞句を登録するにはさまざまな工夫が要る
- 2) 主動詞の前に副詞が入る場合は、別途、新しく登録しなければならない
- 3) 主動詞が必須格を採るシテ形接続を辞書登録できない

などの諸問題の解決を挙げることができる。

翻訳メモリへの登録

今後の実験課題ではあるが、必須格を採る主動詞を有するシテ形接続を含む和文の英訳については、翻訳メモリを活用する。作成データのうち1570件について、名詞変数のみを残したパターンを作成したが、このうち960件は必須格を採る主動詞を有するシテ形接続である。この実験については、別途、報告する。

## 6. おわりに

1 第3項で言及した CREST データ分析によれば、心理表現に関わるシテ形接続の動詞接続句の意味的な論理構成に対応して典型的な英語文型がそれぞれあることが判明した。和文の主動詞が必須格を採るシテ形接続の場合についての分析は、今後の課題である。

2 心理を表す動詞接続句に対応する英語文型には、定型化したパターンと、非定型パターンとを認めることができる。日英間で統語構造の対応をとらうる定型パターンでは、翻訳手法の定式化が可能である。一方、統語構造の対応をとるのが難しい非定型パターンとは、翻訳手法が個別的である場合や、慣用表現の場合である。典型の多くは定型であるが、例外もある。

3 第4項に示した実験結果によれば、日英機械翻訳では、その多くは and 等位接続の文型を採用して訳文を生成する傾向にある。しかし、本稿で報告した CRET データからは、and 等位接続は僅少という顕著な傾向が確認された。シテ形接続の意味解析がされて、理由節へ訳し分けられている場合もあるが(Web1)、意味に応じて適切な文型を選択することは、極めて困難である。この課題解決に向けては、本稿で報告した動詞接続句を、辞書登録データとして採用することで、意味的にも文型的にも適訳を生成する具体的手法の開発が求められる。

4 上記実験結果によれば、用言接続句を一語の複合動詞と見なして辞書に登録することで、翻訳英文が、単純・簡潔な英文へと改善される、ことを報告した。動詞接続句が、英語では一語の動詞 + アルファで表される場合、複合動詞の一種として見なして辞書に登録することができた。さらに、和文の主動詞が必須格を採るなどの他の諸形態について、辞書登録手法の開発が求められる。すでに和文の主動詞が必須格を採るシテ形接続を辞書に登録できる試み[7]があるが、こうした拡張性の大きい機構をユーザ辞書インターフェースに組み込むことが求められる。

5 本稿では、心理表現について分析したが、それ以外のシテ形接続について、そのシテ形接続の意味的な論理構成についての分析は課題として残されている。文型を網羅しての分析によって、普遍相の把握を確定することが今後の課題である。約言すれば、シテ形接続という表現形式に関する、意味類型[8]の体系を揃えることである。

6 意味の表現形式には類型および典型があり、シテ形接続の意味的論理構成に応じて英語の文型を選ぶことが可能である。本報告では、この意味類型実証の端初が得られたといえるだろう。

本研究は、戦略的基礎研究事業「セマンティックタイポロジーによる言語の等価変換と生成」の研究の一環であり、科学技術振興事業団の支援を受けている。論議を頂いた研究プロジェクト各位に、データ作成作業にご協力を頂いた衛藤純司氏(ランゲージウェア)・荻野孝野氏(JSA)に、また機械翻訳実験に協力を頂いた(株)クロスランゲージ社に多謝申し上げる。

## 文 献

- [1] 仁田義義雄, “シテ形接続をめぐって,” 複文の研究(上), pp.87-126, くろしお出版, 東京, 1995.
- [2] 佐良木昌, 新田義彦, “シテ形接続の意味分類体系試案,” 電子情報通信学会全国大会講演論文集 A-13-3, March.2005
- [3] 佐良木昌, “感情論理と感情表現,” 信学技報, vol.106, no.486, TL2006-63, pp. 37-41, Jan.2007.
- [4] Adele, E. Goldberg, 構文文法論 英語構文への認知的アプローチ, 河上誓作ら(訳) 研究社, 東京, 2001.
- [5] G.リーチ, Y.スヴァルトヴィック, “気分、感情、および態度,” 現代英語文法, pp.226-241, 池上恵子(訳), 紀伊國屋書店, 東京, 1998
- [6] 佐良木昌, “シテ形接続と英語文型との対照,” 英語表現学会第35回全国大会, June.2006.
- [7] 柴田勝征, “「～して～する」の英訳文生成パターンの分類 - 拡張された複合動詞句の活用 - ” 言語処理学会第13回全国大会, March. 2007
- [8] 有田潤, ドイツ語講座, 南江堂, 東京, 1987.
- [9] R. Quirk, S. Greenbaum, G. Leech, J. Svartvik, A Comprehensive Grammar Of The English Language, Longman, Essex, 1991.

表2 1パターン文種別頻度

文種	件数	割合
単文	48	78.7%
複文	7	11.5%
分詞構文	4	6.6%
等位接続	2	3.3%
総計	61	100.0%

表3 2パターン文種別頻度

文種	件数	割合
単文	10	83.3%
等位接続	2	16.7%
総計	12	100.0%