周利・玉岡賀津雄の発表(2020年3月14日,名古屋大学にて)

日中対訳コーパスにおける対訳関係の共起頻度指標の吟味と分析法の提案 ---「てくる」の日中対訳の共起頻度を例に---

1. はじめに

現代日本小説100篇をテキストデータにした李光赫編集の『日中文学作品対訳コーパス』や北京日本学研究センターの徐一平が研究代表として編集した『日中対訳コーパス』など、2言語間対訳コーパスを使用した対照研究が行われるようになってきた。李・趙(2018)は、タラ条件文についての中日対訳の関係を、1つのコーパス内の共起頻度の指標であるTスコアとMIスコアを、対訳コーパスに適用して、これらの2つの指標で散布図に描いて日中対訳表現の類似性を検討した。そこで、2言語間の「てくる」の日中対訳の共起頻度(周,2019)を例に、対訳共起頻度指標を計算して、李・趙(2018)のアプローチの是非を検討し、最適な分析法を提案する。

2. 周(2019)の「てくる」をめぐる日中対訳頻度研究

周(2019)は、李が編集した日中対訳コーパスを用い、「てくる」に対応する中国語表現について、A類:V来(単純方向補語)、B類:V(x)来(複合方向補語)、C類:来(趨向動詞)、D類:V進などの(来以外の単純方向補語)、E類:方向補語の伴わずかつ「来/去」でない類、F類:去(趨向動詞)、G類:V(x)去(単純/複合方向補語)の7つに分類した。さらに、中国語の訳の「来」を含む「A類/B類/C類」を「対応類」、中国語の「去」を含む「F類/G類」を「逆対応類」、中国語の「来/去」とも含まない「D類/F類」を「非対応類」とした。また、「移動方向性」と「同時移動」には、「起点/通過点」「着点」「方向」の素性([±])が「てくる」移動構文に明示されるか否かによって、中国語の訳文が異なってくるとした。366例の「てくる」移動構文の対訳頻度を表1に示した。周(2019)は、李・趙(2018)を参考にして、TスコアとMIスコアの散布図から「てくる」の中国語訳の傾向を考察した。

表 1:「てくる」移動構文の下位構文パターンと中国語訳の共起頻度 (9×7のセル, N=366)

日本語		中国語 -	対応類			非対応類		逆対応類	
口平品			A類	B類	C類	D類	E類	F類	G類
移動の 方向性	(1) [-起点/通過点] [-着点]	[-方向]	33	25	12	3	10	0	2
	(2) [+起点/通過点]		11	12	0	0	5	0	0
	(3) [+着点]		0	2	17	60	5	0	0
	(4) [+方向]		6	5	0	2	22	0	0
同時移動	(5) [-起点/通過点] [-着点]	[-方向]	34	18	0	1	2	0	0
	(6) [+起点/通過点]		5	10	0	0	2	0	0
	(7) [+着点]		2	2	1	13	0	0	0
	(8) [+方向]		3	3	0	0	4	0	1
継起移動	(9) 継起移動		3	0	0	2	4	21	3

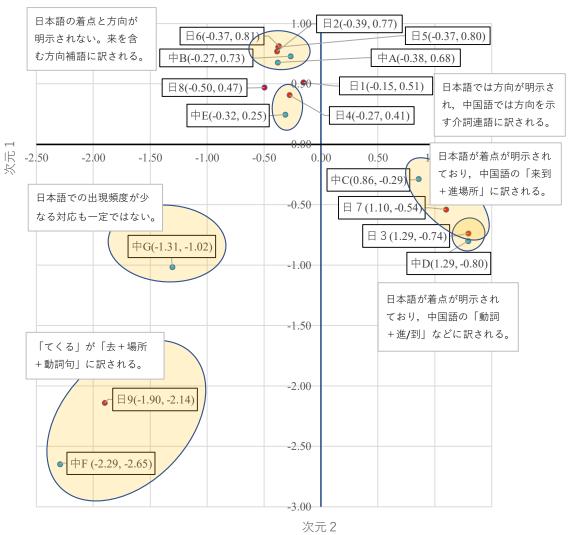
3. 対訳頻度の指標化

表 1 の対訳共起頻度を, (1)共起頻度, (2)共起頻度の対数変換, (3)BiTスコア, (4)BiMIスコア, (5)BiMI³スコア, (6)BiDice係数, (7)BiDice log Freq, (8)BilogDiceの 8 つの指標で示した (注: 2つのコーパスの対訳の共起頻度を計算するために, Tスコアの中心語頻度はFxで, またNはコーパ

スの語彙総数ではなく検索対象語総数として計算して、従来の指標を踏襲した。Biを指標名に追加)。9つの日本語、7つの中国語の63セルの共起頻度から、実際に頻度が存在する37セルから8つの指標を計算した。周(2019)は、李・趙(2018)と同じアプローチで、「てくる」の日中対訳の関係を、BiTスコアと8iMIスコアの2つの指標で示して考察している。しかし、両変数のピアソンの積率相関係数が非常に高く(N=37, r=.91, p<.001)、ほぼ同じ指標であることが分った(2: 発表では指標の数式など詳細も報告する)。類似した2つの指標の散布図から対訳関係を考察することの意味はないので、このアプローチは妥当とはえいないであろう。

4. 対訳頻度を使ったコレスポンデンス分析の提案と考察

表1の共起頻度について、共起頻度の存在する37のセルについて、距離測度をカイ二乗で測定してコレスポンデンス分析を行った。イナーシャの寄与率で、次元1が41.82%、次元2が39.94%で、合計81.76%という非常に高い説明力を示した。2次元散布図で結果を示した。日中対訳のもともとの共起頻度から両者の関係を示すことができた。このアプローチは2つの言語の対訳コーパスの共起頻度から2言語間の関係を示すのに有効なようである(提案)。



参考文献 余白がないので割愛