日中対訳コーパスの共起頻度に基づいた対訳関係解析[1]
—「てくる」の日中対訳の共起頻度を例に—

周　利[2]
李　光赫[3]
玉岡　賀津雄[4]

DOI: 10.18999/stul.36.25

要約　『現代小説 100 冊日中対訳コーパス』[5](李光赫 編)は，日本語から中国語へ翻訳した対訳テキストを収めたコーパスである。このようなコーパスは，対訳コーパス(bilingual corpus)と呼ばれる。李光赫・趙海城（2018, 2020）および周利（2019, 2021）は，この日中対訳コーパスから得られた 2 言語間の共起頻度を使って，*T* スコアと *MI* スコアを計算した。そして，両スコアを散布図に描いて，日中の対訳関係を分類した。ただし，対訳の場合はコーパスが 2 つ別々に存在するため，本研究では，それぞれ *BiT* スコアと *BiMI* スコアと呼んで再定義した。さらに，周利（2019, 2011）の日中対訳コーパスの共起頻度データから，「てくる」の下位構文パターンと中国語訳をコレスポンデンス分析で解析して，5 つのパターンを得た。この解析法は，対訳コーパスで得られた共起頻度をそのまま使って両言語の関係を詳細に記述するのに有効である。この解析法は，共起頻度を基にして，日中の主要な対訳関係を選出するのに有効である。

キーワード　対訳コーパス，*T* スコア，*MI* スコア，*BiT* スコア，*BiMI* スコア

1 Title: Analysis for translational relations on frequencies of co-occurrences in Japanese-Chinese bilingual corpus
2 Author: ZHOU, Li (Graduate Student, Osaka University, Japan) zhouli690121@gmail.com
3 LI, Guang He (Dalian University of Technology, China) liguanghe5588@outlook.com
4 TAMAOKA, Katsuo (Hunan University, China, and Nagoya University, Japan) ktamaoka@gc4.so-net.ne.jp
5 このコーパスは未公開で，大連理工大学の李光赫先生に連絡し，許可を得れば使用することができる。

1. はじめに

対訳コーパス（bilingual corpus）は，異なる言語の文と文が対訳の形でまとめられたコーパスのことである。中日対訳コーパスは，北京日本学研究センターの徐一平先生を研究代表とする中国社会科学基金研究プロジェクト「中日対訳コーパスの構築とその応用研究」の援助を受けて開発され，2003 年に第1版が公開された。日中対訳コーパスの第1版は約 2 千万字のサイズである。なお，徐一平の中日対訳コーパスは，2022 年 9 月現在，アスセスができないようである。また，大連理工大学の李光赫先生が，『現代小説 100 冊日中対訳コーパス』を作っている。このコーパスは，現代日本語で書かれた小説100篇とその中国語訳をテキストにして検索できるようにしたものである。李光赫先生の日中対訳コーパスは未公開になっている。2022年9月18日現在，著作権などの問題があり，個人的に使用する以外に，一般公開された中日あるいは日中対訳コーパスはないようである。

日中の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。たとえば，李光赫・趙海城（2018）は，「タラ」条件文についての日中対訳の関係を，共起頻度の指標である *T* スコアと *MI* スコアの 2 つの指標（詳細は，石川, 2006）で散布図に描いて日中対訳表現を検討した。周利（2019）も同様のアプローチを踏襲している。本研究では，第 1 に，李光赫・趙海城（2018）と周利（2019）の 2 つの研究データを使って，1 つのコーパス内での共起頻度の選択基準に使われた *T* スコアと *MI* スコアを，日中対訳の日本語と中国語の 2 つのコーパス間の共起頻度から対訳関係の解明にどのように適用しているかを検討した。第2に，周利（2019）の「てくる」の日中対訳の共起頻度にコレスポンデンス分析を適用することを試みた。第 3 に，コレスポンデンス分析を使って，周利（2019）の「てくる」の日中対訳の共起頻度を基にして，対訳関係を考察した。

2. 日中対訳コーパスと先行研究

2.1 『現代小説 100 冊日中対訳コーパス』

現代日本語小説100篇を中国語訳と対応させた『現代小説100冊日中対訳コーパス』（李光赫 編）は，現代日本語がどのように中国語に訳されているかを考察する上で重要な対訳コーパスである。このコーパスには，森村誠一の『腐蝕の構造』，西村京太郎の『札幌着 23 時 25 分』，又吉直樹の『火花』，村上春樹の『羊をめぐる冒険』，赤川次郎の『三毛猫ホーム

ズシリーズ』の『三毛猫ホームズの怪談』および『三毛猫ホームズのびっくり箱』，東野圭吾の『白夜行』および『秘密』，田中芳樹の『銀河英雄伝説』（黎明篇，風雲篇など）などの新しい現代小説が含まれている。

2.2 李光赫・趙海城（2018）の「タラ」条件文の日中対訳研究

李光赫・趙海城（2018）は，『現代小説 100 冊日中対訳コーパス』を使って「タラ」条件文の日中対訳関係を調べるため，日本語の「タラ」の 6 種類の意味用法に対応する 15 種類の中国語表現の対訳で，2 つのコーパスの共起頻度（李光赫・趙海城, 2018, p.23, 表 1）を計算した。そして，90 のセルの内（6×15）で，49 セルを埋める共起頻度をみいだした。これは，日本語の 6 種類の意味用法が，49 の中国語に訳されることを示している。共起頻度は 1 から 126 の範囲で分布し，平均（*M*）が 10.20 で，標準偏差（*SD*）が 19.75 であった。*T* スコアは，もともとの 1 つのコーパス内での共起指標で，以下の計算式で示される。

$$T=\frac{fxy-\frac{fyfB}{N}}{\sqrt{fxy}} \quad \cdots\cdots\cdots\cdots (1)$$

f_y は中心語頻度と呼ばれ，共起表現の中心となる語の頻度である。たとえば，「ぽい」（「っぽい」を含む）の場合，「男っぽい」「子供っぽい」などを検索する際の中心となる「ぽい」の頻度である。*fB* は，中心となる語と共起する名詞の共起語頻度である。f_{xy} は，中心語と共起語の共起頻度である。たとえば，「男っぽい」であれば，「っぽい」が中心語で，「男」が共起語になる。両者が共起して出現頻度が計算できる。*N* は，コーパスの総語数である。たとえば，国立国語研究所の『現代日本語書き言葉均衡コーパス』（BCCWJ）であれば，104,911,460 語という大きな数値である。*T* スコアは，*T* 分布に従って絶対値が 1.96 で，5%の有意水準となるので，有意に共起する組み合わせであると判断される。

一方，李光赫・趙海城（2018）の計算は日中対訳コーパスを使っているので，中心語も共起語頻度も存在しない。その代わり，日本語の「タラ」条件文の表現頻度が f_x となり，f_y が中国語に訳した場合の表現の頻度になる。f_{xy} は，日本語の 6 種類の意味用法に対応する 15 種類の中国語表現の対訳の共起頻度である。*N* は，両者の総合共起頻度の 500 であり，これが従来の *T* スコアと大きく異なっている。違いを示すために，数式（2）では N_{xy} とし，日中

両言語の対訳頻度の合計とした。2 言語の対訳コーパスなので，Bi（bilingual）をつけて *BiT* スコアと呼ぶことにする。この式は以下のようになる。

$$BiT=\frac{fxy-\frac{fxfy}{Nxy}}{\sqrt{fxy}} \quad \cdots\cdots\cdots\cdots (2)$$

李光赫・趙海城（2018）は，日中で 2 つのコーパスがあるとして，2 倍して N_{xy}=1,000 として計算している。もともとの *T* スコアでは，数式（1）の *N* がコーパスの総頻度であるため，*BiT* の N_{xy} よりも遥かに大きい数値が入ることになる。本研究では，*N*=500 として，李光赫・趙海城（2018）の *BiT* スコアを数式（2）で計算すると，49 の日中対訳関係（*N*=49）で，*M*=0.10, *SD*=1.79，歪度=−0.73，尖度=1.02 となった。特定の頻度が多く，狭い分布である。なお，N_{xy} を 2 倍して 1,000 として計算しても，ピアソンの積率相関係数は，*r*＝0.97（*N*=49, *p*<.001）となり，両者はほぼ同じ数値になることがわかる。2 つのコーパスだからということで 2 倍しても実質的な違いはない。*BiT* スコアは，最小が−5.64 から最大は 3.07 であった。21 の対訳関係において負の値が得られ，28 の対訳関係が正の値となった。*BiT* スコアでは，N_{xy} が小さいので，*t*検定で 5%の有意水準として使われる絶対値 1.96 という境界値はあまり有効ではないであろう。李光赫・趙海城（2018）では，正の値（0 以上）の対訳関係を意味のある頻度であるとして分析の対象としている。

T スコアは，語の頻度情報を重視するために高頻度語を共起関係が強いと評価する傾向がある。そこで，頻度は低いが，特殊な結びつきをしている共起表現を検出するために *MI* スコアが考案された。*MI* スコアは，相互情報量ともいわれ，*T* スコアと同様に 1 つのコーパス内での共起頻度の有用性を判定する指標であり，以下の式で得られる。

$$MI=log_2 \frac{fxyN}{fxfy} \quad \cdots\cdots\cdots\cdots (3)$$

李光赫・趙海城（2018）は，この数式（3）を 2 つのコーパスからなる日中対訳コーパスにも適用した。数式そのものは一見同じようにみえるが，*T* スコアと同様に *N* の扱いが異なる。やはり，日中両言語の対訳頻度の合計を N_{xy} とした。もともとの *MI* スコアと区別するために，

ここではBiをつけて*BiMI*スコアと呼ぶことにする。対数(*log*)の底は2であるが，最近の対数計算では自然対数を使うことが多い。なお，自然対数の底は *e* で示され，数値は *e*=2.7182…，となる。円周率(円の直径に対する円周の長さの比率)が 3.1415…と無限に続くように，*e* も無限に続く。対数を発明したスコットランド生まれのジョン・ネイピア(John Napier)の名前をとって，ネイピア数(Napier's constant)とも呼ばれる。

$$BiMI=log_2 \frac{fxyNxy}{fxfy} \quad \cdots\cdots\cdots\cdots (4)$$

*T*スコアの場合と同様に，*MI*スコアと*BiMI*スコアは，N_{xy}の値を両者の総合頻度の500(2倍した1,000でも同じ)で計算した。李光赫・趙海城(2018)の共起頻度データに*BiMI*スコアの数式(4)を当てはめると，f_x は日本語の「タラ」条件文の表現頻度，f_y は中国語に訳した場合の表現頻度，f_{xy}は日本語の「タラ」の6種類の意味用法に対応する15種類の中国語表現の対訳の共起頻度となる。*MI* スコアを 49 の日中対訳関係について計算すると，最小が −2.73 で，最大が 4.24 となった。また，分布は，*N*=49, *M*=0.49, *SD*=1.53, 歪度=0.62, 尖度=0.12となった。ちなみに，N_{xy}=1,000とした場合と N_{xy} =500とした場合のピアソンの積率相関係数は，*r*=1.00(*N*=49, *p*<.001)となり，N_{xy} の値を変えた両方の*BiMI*スコアの数値がまったく同じ値であることを示している。

*BiT*スコアと*BiMI*スコアは，共に稀な共起表現を削除するための基準として作られている。つまり，両スコアの目的が同じであるため，両スコアの相関が高くなることが予想される。ピアソンの積率相関係数を計算すると，*N*=49, *r*=0.82, *p*<.001 で，非常に高かった。李光赫・趙海城(2018, p.25, 図1)は，両スコアを0から5までで変化する散布図に描いている。そして，0以上の部分について，Y軸の*BiT*スコアを約1.5，X軸の*BiMI*スコアを約2.0の線で4つのブロックに分けている。0以上であっても両スコアが低いブロックは考慮せず，両スコアが高い対訳頻度をAブロックとして「相互的対応関係」，*BiT*スコアが高く*BiMI*が低い場合をBブロックとして「中国語訳傾向」，*BiT*スコアが低く*BiMI*が高い場合をCブロックとして「タラの意味情報を含む中国語形式」として，ブロックごとに議論している。ある程度，稀な共起頻度の対訳については考慮せず，主要なものだけを取り出して議論するというアプローチである。

2.3 周利（2019）の「てくる」構文の日中対訳研究

周利（2019）は，『現代小説 100 冊日中対訳コーパス』を使って，表1に示したように日中対訳の共起頻度を計算した。周利（2019）も，李光赫・趙海城（2018）を参考にして，*BiT* スコアと *BiMI* スコアの散布図から「てくる」構文の中国語訳の傾向を考察している。周利（2019）は，「てくる」構文に対応する中国語表現について，下位構文パターンを日本語は 9 種類，対応する中国語表現は 7 種類に分類した。具体的には，

A 類: “V（動詞）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 进，到，回”などの（“来”以外の単純方向補語）
E 類: 介詞連語
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/E類」を「非対応類」とした。また，「移動の方向性」と「同時移動」には，「起点/通過点・着点・方向」が「てくる」構文に明示されるか否かによって，中国語訳が異なってくるとした。「てくる」構文の特徴から 9 種類（以下，日 1 から日 9 で示す）に分類し，それに対応する中国語表現を 7 種類（以下，中 A から中 G で示す）に分類した。

表 1 「てくる」構文の下位パターンと中国語訳の共起頻度

日本語 ＼ 中国語		対応類			非対応類		逆対応類	
		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

366 例の「てくる」構文の対訳頻度は，表 1 に示した。全体が 63 のセル（9×7）で，そのうちの 37 セルに共起頻度がみられた。頻度は，1 から 59 の範囲に分布し，平均が 24.68 で，標準偏差が 18.47 であった。*BiT* スコアは，総頻度の N_{xy} が 366 の 37 セルの日中対訳関係で，最小値が−11.08，最大値が 5.35 になった。分布は，*N*=37，*M*=−0.55，*SD*=3.75，歪度=−1.44，尖度=2.33 となった。*N* を 2 倍して 732 として計算しても，ピアソンの積率相関係数は，r＝0.97（n=37，p<.001）となりほぼ同じ値を示した。37 セルの日中対訳関係で *BiMI* スコアを計算すると，最小値は−3.14 で，最大値は 3.47 であった。分布は，*N*=37，*M*=0.04，*SD*=1.62，歪度=-0.26，尖度=−0.15 となった。なお，*N* を 2 倍して 604 として計算すると，ピアソンの積率相関係数は r＝1.00（n=37，p<.001）となる。つまり，*BiT* スコアはわずかではあるが N_{xy} によって変化するが，*BiMI* スコアは N_{xy} では変わらないことがわかる。

2.4 *BiT* スコアと *BiMI* スコアによる対訳関係の考察

李光赫・趙海城（2018）は，数式（2）の *BiT* スコアと数式（4）の *BiMI* スコアを考案して，日中対訳の関係を考察した。対訳コーパスに応用した *BiT* スコアと *BiMI* スコアは，コーパスの総語数である N_{xy} がもともとの *T* スコアと *MI* スコアとは異なるが，両者のスコアの特性を温存している。周利（2019）のデータを *BiT* スコアと *BiIM* スコアの値で散布図に描くと，図1にようになった。

BiT スコアと *BiMI* スコアは，近似直線（R^2=0.819，この分析は小数第3位まで示す）が Y=2.108X－0.380 となり，ピアソンの積率相関係数は非常に高く（*N*=37，r=0.905，p<.001），両スコアが類似した指標であることがわかる。ここで，李光赫・趙海城（2018）および周利（2019）の研究に従い，0以下の対訳関係を頻度が小さいので排除する。さらに，両スコアが1以下の対訳を排除する。その結果，黒丸の●で示された 21 の対訳関係は排除される。そして，図1の白丸の○で示された 16 の対訳関係が残った。これらは灰色で塗られた図1の部分に所属する対訳関係である。図1から，この灰色の範囲の対訳関係はある程度分散していることがわかる。これらが「てくる」の主要な日中対訳と考えられる。このように，*BiT* スコアと *BiMI* スコアは，『現代小説 100 冊日中対訳コーパス』を使って，特定の主要な対訳関係を明瞭に示すのに秀でた指標である。なお，李光赫・趙海城（2018）および周利（2019）では，より厳しく 2 に近い基準で区切っている。

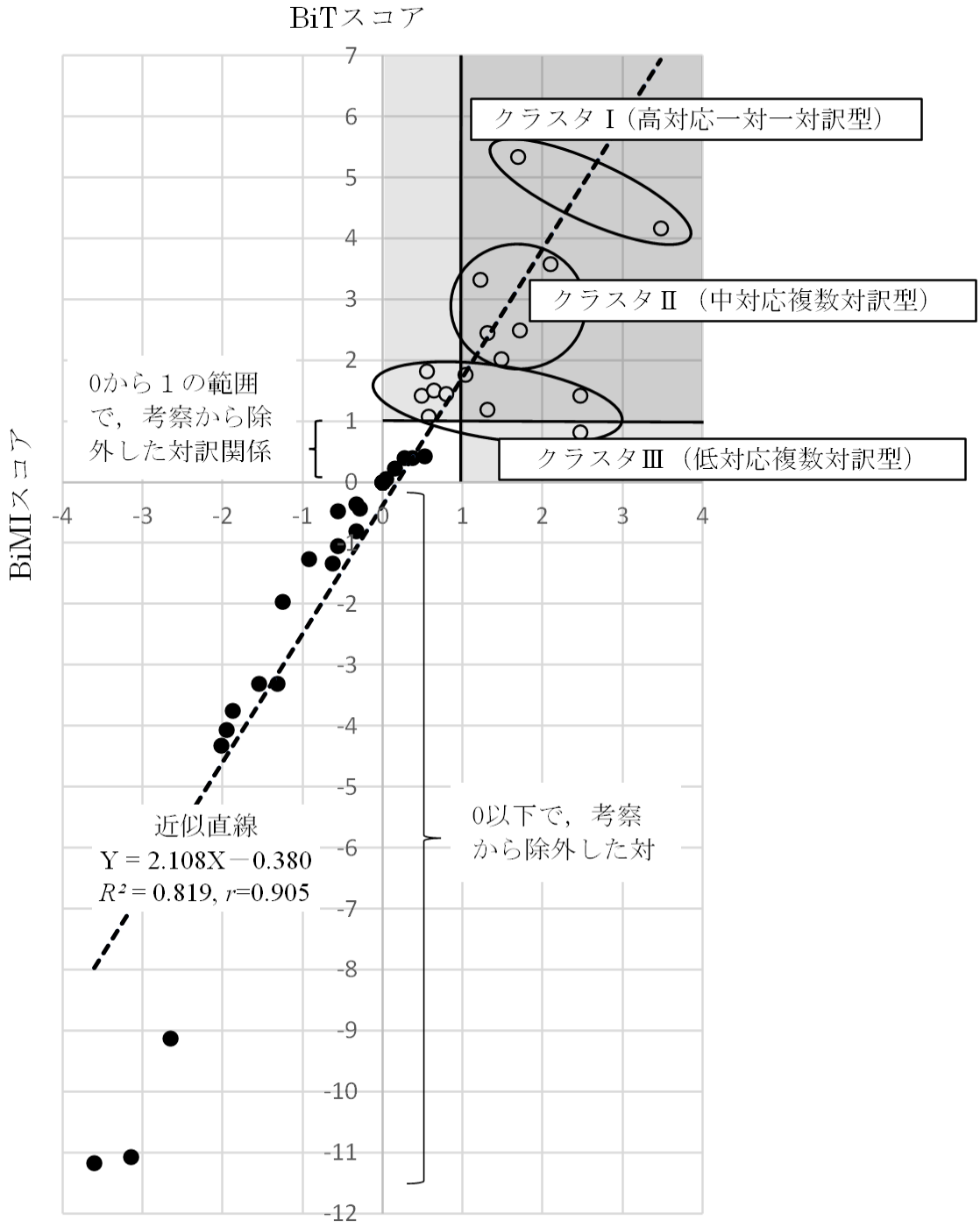

図1　*BiT* スコアと *BiMI* スコアの散布図とクラスタ分析の分類

周利（2019）の 16 の対訳関係の頻度を使って，グループ間の距離はウォード法，個々の対訳間の距離はユークリッド距離でクラスタ分析を行った。その結果は，図 2 に示したように，25 ポイントのスケールの 10 ポイントの位置で 3 つのクラスタに分けられた。念のために，こ

れらの 3 つのクラスタを判別分析で逆に予測すると，交差妥当化での的中率は 93.8%であり，非常に正確に分類されていることがわかった。クラスタ分析で得られた3つのクラスタの対訳関係は，それぞれに異なる特性を持っている。

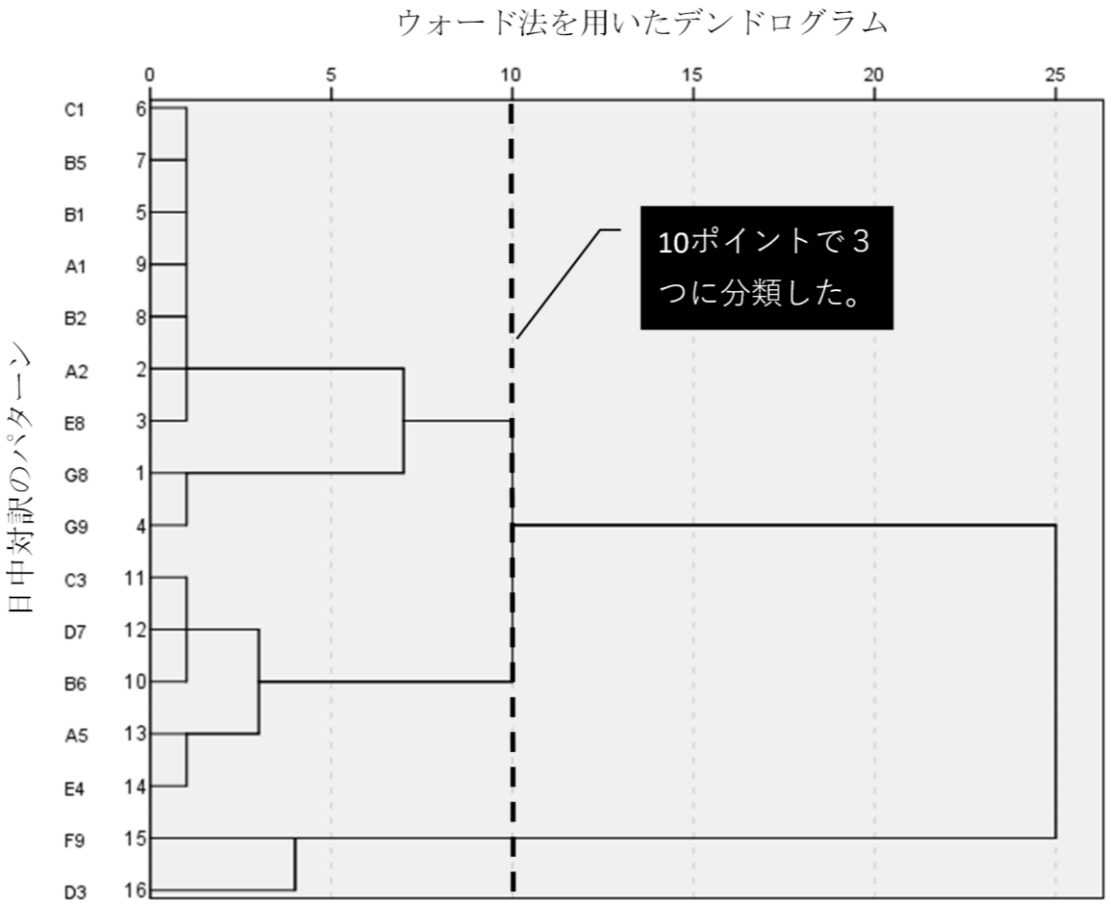

図 2　*BiT* スコアと *BiMI* スコアの数値を使ったデンドログラム

クラスタIは対訳関係が下位構文パターン内で密な「高対応一対一対訳型」の翻訳で，着点が明示される日 3 が中 D 類（60 回）の“V 进，到，”などの“来”以外の「中対応複数対訳型」に，日 9 が中 F（21 回）の趨向動詞の“去”に訳されて“来”に訳されない対訳型である。つまり，高い頻度で中国語の“来”に訳されない関係を示している。また，これ以外の他の対訳がほとんどなく，ほぼ一対一の関係であることも，対訳型の特徴である。クラスタIIは「中対応複数対訳型」で，日中対訳関係が下位構文パターン内で中くらいの頻度で，複数の高頻度の対訳が同時に存在する。クラスタIIIは「低対応複数対訳型」で，低対応関係であり，同

時に共起頻度の高い対訳関係が複数存在する傾向がある。以上のように，*BiT* スコアと *BiMI* スコアは，対訳頻度を基準として対訳関係を分類するのに適したアプローチであることが窺える。

3. 対訳頻度を使ったコレスポンデンス分析の提案と考察

BiT スコアと *BiMI* スコアの散布図による分類は，「てくる」の 9 種類の下位構文と 7 つの中国語の対訳頻度から主要な対訳関係の基準を決めるというアプローチである。しかしながら，もともと対訳コーパスで得られる頻度そのものが少ない。そのためすべての対訳関係の頻度をそのまま活かすために，すべての対訳（共起）頻度をそのまま使用するほうがよいのではないかと思われる。そこで，すべての日中対訳を視覚的に示して，日本語の「てくる」構文がどのように中国語に訳されるかを詳しく考察するために，コレスポンデンス分析を使用する。表1のすべての日中対訳の共起頻度データをコレスポンデンス分析で解析して，分類を試みてみた。なお，李光赫・趙海城（2020）もコレスポンデンス分析を使用している。

ちなみにコレスポンデンス分析は，複数の質的な観点から類似したデータをまとめていくために行われる多変量解析である。本研究のようなクロス集計などの行と列からなる頻度データにおける項目間の関係を視覚的に把握するために 2 次元の散布図を描いて視覚的に示す統計解析法である交点の 0 に近づくほど，このデータ内での平均的な共起関係であることを示す。

「てくる」構文と中国語の“来”は共に話者の視点を含み，人や物あるいは行為などが話者の領域に向かう移動を示す。しかし，対訳の例から，「てくる」構文は中国語の“V 来”あるいは“V（x）来”だけではなく，多様な訳がみられる。そこで，多様性をそのまま考察するために，表1の「てくる」の対訳頻度をそのまま使って分類的な分析を行う。具体的には，表1の「てくる」の日中対訳で共起頻度が存在する 37 のセルについて，距離測度をカイ二乗で測定して，対照的正規化によるコレスポンデンス分析を行った。対訳頻度はノンパラメトリックデータであるため，カイ二乗で距離を測定して分類するのが最適であろう。イナーシャの寄与率では，次元 1 が 41.82%で，次元 2 が 39.94%となり，合計 81.76%という高い説明力を示した。次元1をＹ軸，次元2をＸ軸として，図3のような散布図を描いた。全体から考察して，次元1は「方向の明示」で，方向が明示されている対訳から明示されていない対訳への変化を示すと考えられる。次元 2 は「着点の明示」で，着点が明示されている対訳から明示さ

れていない対訳への変化を示すと考えられる。

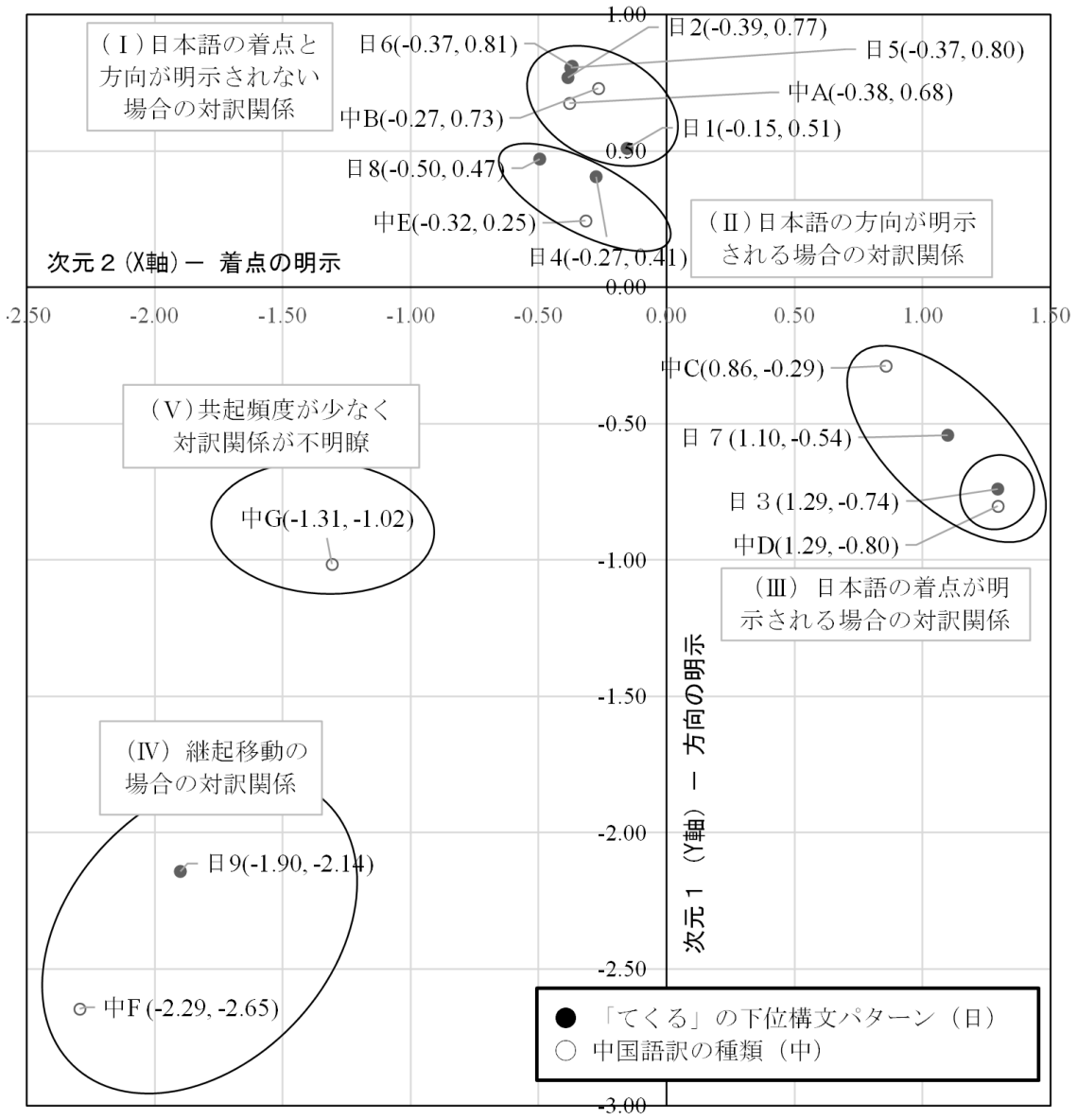

図３ コレスポンデンス分析による日中対訳関係の散布図

図３に示したコレスポンデンス分析の結果に基づく「てくる」の９種類の下位構文（日１から日 9）と７つの中国語の対訳（中 A から中 G）の関係は，散布図の位置関係から，さらに５つの傾向がみてとれる。それらは，図３の楕円の分類である。

(I) 日本語の着点と方向が明示されない場合の対訳関係

(II) 日本語の方向が明示される場合の対訳関係

(III) 日本語の着点が明示される場合の対訳関係

(IV) 継起移動の場合の対訳関係

(V) 共起頻度が少なく対訳関係が不明瞭

の 5 つである。以下，それぞれについて記述的に考察する。

4. 日中の対訳関係のパターン

4.1 日本語の着点と方向が明示されない場合の対訳関係（I）

図 3 の「日本語の着点と方向が明示されない場合の対訳関係（I）」では，日本語の着点と方向が明示されない日 2，日 5，日 6 は，中 A と中 B との関連が強い。言い換えれば，着点と方向が明示されない「移動の方向性」と「同時移動」の用法である「てくる」構文は，中国語の“来”を含む方向補語である単純方向補語の“V 来”あるいは複合方向補語の“V（x）来”に訳される傾向がみられる。ただし，日1のように，対応類である中 A と中 B に訳されるのみならず，非対応類の中国語の介詞連語に訳されることもある。さらに，日本語の着点と方向が明示されない場合の対訳関係（I）は，散布図の原点に近く，「てくる」の典型的な対訳関係である。

4.2 日本語の方向が明示される場合の対訳関係（II）

図 3 の「日本語の方向が明示される場合の対訳関係（II）」では，「移動の方向性」で方向が明示されている日 4 は，中 E と強い関係性がみられる。「同時移動」で方向が明示されている日 8 のような場合は，中 A，中 B，中 E の中国語に訳される。つまり，「日本語の方向が明示される場合の対訳関係（II）」において，日本語の方向が明示される「移動の方向性」と「同時移動」の用法には，「てくる」構文が中国語の介詞連語に訳されるという偏りがある。もちろん中国語の翻訳には介詞連語以外に，中国語の“来”を含む方向補語に訳すこともある。しかし，全体からみると，「てくる」構文に方向が明示されている場合と中国語の「方向や受け手などを示す介詞連語」との対訳関係が強いといえるであろう。なおかつ，この場合の対訳関係も，図 3 の散布図の原点に近いことから，典型的な対訳関係であると思われる。

4.3 日本語の着点が明示される場合の対訳関係（III）

図 3 の「日本語の着点が明示される場合の対訳関係（III）」では，日本語の着点が明示される日 3 と日 7 は，中 D または中 C に対応している。要するに，日本語の着点が明示される「移動の方向性」と「同時移動」は，中 D の「“V 进，到，回”など＋場所名詞」もしくは中 C の「“来到”＋場所名詞」に訳される傾向がある。特に，「移動の方向性」の日 3 は，中 D と強く関連している。さらに，「移動の方向性」で着点が明示されている場合は，中国語では，「“V 到，进，回，上”などの（“来”以外の単純方向補語）」で表現される。また，日本語の着点が明示される「てくる」構文は，中 C の趨向動詞の「“来到”＋場所名詞」に訳されることもある。「日本語の着点が明示される場合の対訳関係（III）」では，稀に「“来到”＋場所名詞」で示されることがあるものの，原則として，着点が明示される「移動の方向性」と「同時移動」の用法の「てくる」構文は，“来”以外の単純方向補語である着点を示す中 D の“V 进，到，回”などに訳される。

4.4 継起移動の場合の対訳関係（IV）

日 9 の実例で描写する移動は移動主体である話者がある場所に行って，そこで実質動詞の動作をしてから，また帰ってくるという移動であり，それは往復的な「継起移動」とも呼ばれている。図 3 の「継起移動の場合の対訳関係（IV）」から，日 9 は，中 F と強い関係を持つことがわかる。しかも，「移動の方向性」・「同時移動」と異なっており，着点と方向が明示されるか否かは，“来”の使用とは関係が薄い。このことは，この場合の「てくる」構文が中国語の「“去”＋（場所名詞）＋動詞句」と訳される傾向があることを示している。これは，加藤（2006）が指摘した，日本語が「戻る過程」をより強く意識するのに対し，中国語が「行く過程」を強く意識することにつながっていると考えられる。最後に、この場合の対訳関係は，図 3 の原点から離れた位置にあることから，偏りの大きい非典型的な対訳関係であることがわかる。

4.5 共起頻度が少なく対訳関係が不明瞭（V）

「共起頻度が少なく対訳関係が不明瞭（V）」では，中国語訳の“V（x）去”（単純/複合方向補語）の中 G が，日本語の「継起移動」の日 9，「移動の方向性」の日 1，「同時移動」の日 8 に対応しているものの，共起頻度は低く，この種の中 G と関連した「てくる」との対応がほとんどみられない。そのため，図 3 では，頻度の低い稀にしかみられない対訳関係であること

がわかる。

4.6 対訳頻度に対するコレスポンデンス分析の有効性

コレスポンデンス分析の結果に基づいて，「てくる」の9種類の下位構文パターンと7種類の中国語表現との対訳関係およびそれらの特性を考察した。また，コレスポンデンス分析であれば，視覚的に全体の対訳パターンを描いてくれるので，図 3 のようにわかりやすく分類を提示することができる。さらに，0 に近い対訳が典型的な対訳であり，0 から遠い対訳が非典型的な対訳であることも，対訳パターンを理解するのに役立つ。以上のように，2 言語間で対訳関係の特徴を共起頻度で詳細に検証するには，対照的正規化によるコレスポンデンス分析が有効な解析法であると思われる。

5. まとめ

現代小説 100 篇を集めた日中対訳コーパス『現代小説 100 冊日中対訳コーパス』(李光赫 編)のおかげで，対訳関係を容易に考察できるようになった。検索でみいだされた多数の表現から規則性をみいだすために，李光赫・趙海城（2018）および周利（2019）の先行研究では，1 つのコーパス内の共起表現の有効性を評価するために考案された *T* スコアと *MI* スコアを散布図に描いて分類を試みた。ただし，対訳ではコーパスが 2 つあり，コーパスサイズの定義が異なるため，本研究では *BiT* スコアおよび *BiMI* スコアと呼んだ。*T* スコアは，共起表現の頻度の高さに焦点を当て，共起頻度の低い表現を排除するための指標であり，*MI* スコアは，低頻度でもユニークな共起表現も抽出できるように補正された指標である。*BiT* スコアと *BiMI* スコアは，こうしたもともとの指標の特性を温存しているようである。図1からも推測できるように，頻度を基にして，日中の主要な対訳関係を選出するのに有効であると思われる。

対訳頻度の低い対訳を含み込むためには，すべての頻度を使って詳細を考察するほうがよいと考えられる。そこで，コレスポンデンス分析を試すことにした。実際，図2に示したように，周利（2019）の表1の「てくる」構文の対訳頻度を基にしたコレスポンデンス分析の結果から，日中対訳関係を 5 つのパターンに分類できた。これらの 5 つの対訳関係から，「移動の方向性」と「同時移動」の 2 つの軸で区別することができ，項目間の要素としての起点/通過点，着点，方向のうち，着点または方向が「てくる」移動構文に明示されるか否かによって，

中国語訳が異なってくることがわかる。

具体的には，第1に，着点と方向が明示されない「てくる」構文は，中国語の“来”を含む方向補語に訳される傾向がみられた。第 2 に，日本語の方向が明示される「てくる」構文は中国語の「介詞連語」に訳される傾向がある。第 3 に，着点が明示される「てくる」構文は，着点を示す“V 进，到，回”などの（“来”以外の単純方向補語）で表現する場合がほとんどである。要するに，着点と方向が明示されていない場合，「てくる」は中国語の“来”を含む中国語表現に訳されやすく，着点か方向が明示されている場合，中国語の“来”に訳されにくいといえるであろう。第 4 に，着点と方向が明示されるかどうかにかかわらない「継起移動」においては，話者自身の往復の移動を中国語の「“去”＋（場所）＋動詞句」で表現することが多い。第 5 に，中国語の“去”を含む方向補語の翻訳は出現頻度が少ないため，関連付けの「てくる」構文はほとんど存在しない。

本研究では，コレスポンデンス分析によって，日本語の「てくる」の多様な意味が，中国語でどのようなパターンに翻訳されるかを，図 3 のように視覚的に示すことができた。このように，コレスポンデンス分析は，「てくる」の日中対訳関係を詳細に分析するのに適した手法であるといえよう。

［参考文献］

石川慎一郎 (2006).「言語コーパスからのコロケーション検出の手法―基礎的統計値について―」『統計数理研究所共同研究レポート』190, 1-28.

加藤晴子 (2006).「中日対訳コーパスにみる“来”“去”と「くる」「いく」の対応状況」『応用言語学研究』 8, 87-104.

周利 (2019). 『「てくる」構文の日中対訳傾向とその事態把握』名古屋大学大学院人文学研究科修士論文

周利 (2021).「『てくる』における日中対訳の特徴と中国語を母語とする日本語学習者の使用状況との関係 ： I-JAS 中間言語コーパスに基づいて」『日本語・日本文化研究』 31, 191-205

李光赫・趙海城 (2018).「関数検定から見るタラ条件文の中国語訳ストラテジー研究」『明星国際コミュニケーション研究』 10, 15-28.

李光赫・趙海城 (2020).「タラ条件文の日中対訳体系のビジュアル化：現代小説 100 編での

タラ条件文 1000 例を中心に―」『ことばの科学』 34, 149-166.

Rychlý, Pavel (2008). A lexicographer-friendly association score, Petr Sojka and Aleš Horák (Eds.), *Proceedings of Recent Advances in Slavonic Natural Language Processing RASLAN2008* (pp. 6-9), Masaryk University, Brno, Czech Republic.

周 利

（大阪大学大学院言語文化研究科・博士後期課程・大学院生）

李 光赫

（大連理工大学・外国語学院・副教授）

玉岡 賀津雄

（湖南大学外国語学院・教授，名古屋大学大学院人文学研究科・名誉教授）