

日本小説の英訳本における 翻訳者の文体的特徴

——東野圭吾作品を例に——

The Translators' Style of English Translations of Japanese
Novels: A Corpus-based Study in Keigo Higashino's Novels

大 羽 良

要 旨

本研究では、東野圭吾の4作品と異なる翻訳者によって翻訳されたそれらの英訳からパラレルコーパスを構築することで原作者の文体によるバイアスを抑えつつ、計量言語学的方法を用いて翻訳者に起因する文体の特徴——テキストに残る「指紋」——を特定することを試みた。またパラレルコーパスから翻訳テキストの文体的特徴として得られた言語選択のパターンを、翻訳者個々に起因する文体的特徴と翻訳テキストが普遍的に持つとされる一般的特徴に分類した。翻訳者個々の文体的特徴から生じるテキスト間の差として「高頻度前置詞の使用率」や「カンマの数のばらつき」等、一方翻訳テキストが持つ一般的特徴に、「短文化」や「翻訳テキストでの文の加筆」、「間投詞の加筆」、「語彙の豊富さの平準化」等が分類される可能性が示された。

キーワード

パラレルコーパス, 計量文体論, 翻訳, 英訳小説, 東野圭吾

1. はじめに

Baker (2000) によれば、翻訳者は自身が翻訳したテキストの中に意識的もしくは無意識的に“thumbprint”のようなものを残すが、それを見つけ出すには方法論的に困難を伴うという。翻訳者の“thumbprint”つまり

翻訳者に起因する文体の特徴を、原作者の文体の特徴やソースとなる言語の一般的傾向から切り離す方法はまだ明確ではないので、そのようなバイアスとなる変数を固定して翻訳者の文体の特徴を完全に明らかにすることは極めて困難であることは確かである。しかし、いくつかの変数の変動を可能な限り抑えることによって翻訳者の文体の特徴を特定する試みは可能であると思われる。

本研究では、一人の作家による日本小説と異なる翻訳者によって翻訳された英訳からパラレルコーパスを作成することで、原作者の文体による変数の変動を抑えつつ、計量言語学的方法を用いることで、翻訳者に起因する文体の特徴を特定することを試みる。

Mikhailov and Cooper (2016) で指摘されているように、このようなパラレルコーパスによる分析はまだ十分であるとはいえず、得られた結果は一般化されているとはいえない。本研究から得られる結果は翻訳者の文体の特徴についての一般化への一助となると考える。

2. 先行研究

2.1 コーパスに基づいた文芸翻訳テキストの文体研究

Baker (2000) では、翻訳されたテキストには翻訳者自体の文体が見つかるとされ、文芸翻訳の文体を調査するためにコーパスを用いた計量言語学的方法論の枠組みが提示された。ここでいう翻訳者の文体とは、翻訳されたテキストから見つかるその翻訳者らしい表現であり、他の翻訳者と比較したとき、より特徴的な言葉の使い方や個人的な言語習慣をあらわす。彼女はそれを“thumbprint”，また金 (2021) では、「指紋」や「DNA」のようなのだと表現している。

Baker (2000) ではコーパスの使用によって、特徴的とはいえども一度しかあらわれない言語選択よりもパターン化された言語選択、さらに翻訳者

が意識的に行ったと思われる言語選択だけではなく、翻訳者の意識を超え、かつ読者の潜在意識で認識できるような言語選択のパターンを捉えようとした。彼女はマンチェスター大学の翻訳研究センターで構築された様々な外国語から英語に翻訳されたテキストをコーパス化した The Translational English Corpus を使用し、ポルトガル語・スペイン語・アラビア語の文芸作品の英訳テキストについて、語彙の豊かさ・文の長さ・報告動詞 say の形態的・文法的特徴（活用形のパターンや動詞に続く that 節の that の有無）を記述した。しかし彼女も認めているように、そこで記述された文体の違いはソーステキストやそのジャンル、原作者自身の文体に起因するものか、ソーステキストが書かれた言語に起因するものか、または翻訳者自身の文体なのかを判断することは難しい。

Olohan (2004) のケーススタディでは、翻訳テキストにあらわれる否定の省略形について研究が行われている。二人の翻訳者による翻訳テキストにあらわれる否定省略形の数の違いは、ソーステキストのジャンル、物語の書かれ方、原作者自身の文体に起因するものだと結論付けている。これは（著者が目的とした）翻訳テキストにあらわれる特徴の一つの発見ではあるが、翻訳者自身の文体を特定することを目的とすれば期待した結果が得られたわけではない。

Mikhailov and Cooper (2016) の入門書では、ロシア語とそのフィンランド語訳から構築したパラレルコーパスから、ソーステキストと翻訳テキストの文の長さと言語の豊富さを調査し、翻訳者の文体を記述する研究を紹介している。ソーステキストが同一言語であるため、ソーステキストでの比較を踏まえたのち翻訳テキストの文体が調査できるという点で、Baker (2000) よりも翻訳者の文体を特定しやすくなっている。しかし彼らはソーステキストを時代別に分けサブコーパス化し比較を試みたが、同時代内でも文体の特徴の異なる作家や時代を超えて文体の影響を受けた作家の存

在などのバイアスによって、最終的にソーステキストの作家別・翻訳者別の小さなデータ間での比較にとどまった。彼らも得られた結果に対して一般化することは難しいことを認めている。

Mikhailov and Cooper (2016) のようにソーステキストからのいくつかのバイアスを取り除くことができれば、翻訳テキストの文体の特徴をつかむことは容易くなる。しかし既存の翻訳パラレルコーパスでは、翻訳者の文体の特徴を一般化するためのデータを十分に得ることは難しい。一般に向けて公開するコーパスは著作権の問題をクリアしなければならないなどの問題を抱えているのも一因であると推測される。翻訳者の文体の特徴をつかむためには、非公開にはなるが新たにパラレルコーパスを作成し、翻訳テキストにみられる翻訳者の文体の特徴を特定する試みが考えられる¹⁾。

2.2 翻訳テキストに見られる四つの普遍的特徴

Baker (1996) によれば、翻訳テキストには四つの普遍的特徴がみられるという。それは明示化 (explication)、簡素化 (simplification)、正規化・保守化 (normalisation/conservatism)、そして平準化 (leveling out) である。

明示化とは、翻訳においてソーステキストに暗示される情報を翻訳テキストに表出する行為をさす。計量言語学的には、文の長さについて翻訳テキストがソーステキストよりも長くなったり、統語マーカー（英語における that 節の that など）・接続語・説明のための語句が増加するといわれている。

簡素化とは、翻訳において長文を分割して訳したり、ソーステキストが持つ曖昧さを解消して訳したり、一般的でない位置に置かれた句読点を修正したり、語彙をソーステキストよりも減少させることなどを含む翻訳行為をさす。

正規化・保守化とは、ソーステキストにみられる文法構造・句読点法・コロケーション・定型表現などを翻訳される言語に合わせて変化させるこ

とをいう。ソーステキストにおける実験的な句読点法が翻訳されるときには避けられることもこれに含まれる。

平準化とは、一般的に翻訳されたテキストは極端な計量言語学的特徴（語彙の豊富さ・文の長さ・口語的か文語的かなど）を失い、中庸化する傾向があることをさす。

翻訳テキストには、そこにあらわれる具体的な言語特徴が翻訳者個々に起因する文体の特徴だけでなく上記の普遍的特徴を含んでいる可能性があり、それを特定し、翻訳者による文体の特徴と切り離す試みには大きな意味があるだろう。

3. 研究目的

本稿では、できる限りソーステキストが翻訳テキストにバイアスをもたらさない形でコーパスを構築し、日本小説を英訳するというプロセスにおいて翻訳テキストにあらわれる翻訳者の文体の特徴を計量言語学的手法を用いて特定することを目的とする。また特定した具体的言語的特徴がBaker (1996) に挙げられた普遍的特徴を含んでいるのか、もしくは翻訳者自身の文体なのかどうかを検証する。

4. コーパスデザイン

本研究では、研究目的に沿うコーパス構築のために、日本小説から東野圭吾の4作品を選択した。

東野圭吾は日本国内で最も著名な作家の一人であり、またデビューの1985年より継続的に小説を発表し、ベストセラーとなった小説の数も多い。それらのいくつかの著作は様々な言語に翻訳され、英訳された著作も10作品にのぼる。また Alexander O. Smith が英訳した *The Devotion of Suspect X* は2012年のエドガー賞長編賞にもノミネートされた。日本小説が英訳される

場合、一人の作家の著作を一人の翻訳者が翻訳することが多いが、東野圭吾の著作は、2021年3月において主には前述した Alexander O. Smith によって訳された著作が多いものの（6冊）、他の翻訳者に翻訳された著作も4作品ある。

本研究では、英訳された作品の中から翻訳者の異なる以下の4作品の小説の本文のみをコーパス化し、分析に用いた。使用した作品を表1に示す²⁾。原作の4作品の間には刊行までに最大24年、英訳本の刊行の間には15年のひらきがあるが、本研究ではその期間での文体の時間的変化については考慮しないこととした。

今回は日本語のソーステキストとそれに対応する英訳されたテキストから成る日英対照のパラレルコーパスを作成した。対応は一文対一文を基本としたが、訳文が分割されたり、結合されたりした場合は数文をひとまと

表1 使用した作品名

ソーステキスト		翻訳テキスト		
作品名	刊行年	作品名	刊行年	翻訳者
秘密	1998	Naoko	2004	Kerim Yasar
容疑者Xの献身	2005	The Devotion of Suspect X	2011	Alexander O. Smith
新参者	2009	Newcomer	2018	Giles Murray
ナミヤ雑貨店の奇蹟	2012	The Miracles of the Namiya General Store	2019	Sam Bett

表2 構築したパラレルコーパスの抜粋（『秘密』と Naoko コーパスから抜粋）

No.	ソーステキスト	翻訳テキスト
465	「さあ、そばに来てあげたよ。」	'Okay, I'm right here next to you.
466	何だい？	What is it, sweetie?
467	何でも話さない」	Just tell me whatever you want.'
468	彼は優しくいった。	φ
469	だが藻奈美はすぐには唇を開こうとはせず、じっと彼の顔を見つめた。	Monami didn't answer right away. She simply stared intently at his face.

まりとして対応させている。また対応がない場合は空白とした。例として表2にコーパスの抜粋を示す。

またソーステキストと翻訳テキストはそれぞれ、形態素解析エンジン Mecab (ver. 2.7.0 IPA 辞書使用) と品詞付与ソフトウェア Stanford POS Tagger (ver.4.2.0) を用いて、単語に分割し品詞情報および辞書見出し語 (原形・Lemma) を付与した。なお日本語のソーステキストでは、Mecab に正しく解析させるために固有名詞や旧字体など116語を辞書に加えている。

結果、コーパスの構成は表3のとおりとなった。なお、語数において記号はカウントしていない。文数をみると、ソーステキストよりも翻訳テキストのほうがすべての作品において少なくなっていた。語数も同様に翻訳テキストのほうが少なくなっている。しかし日本語を形態素解析エンジンで単語に分割する場合は、その解析エンジンによって短単位・長単位と分割の長さが異なり (Mecab の場合は短単位となる) それによって結果が変化する可能性もあるので、ソーステキストと翻訳テキストの比較はあまり意味をなさないかもしれない³⁾。

表3 コーパスの構成

作 品 名	語 数		翻訳／ ソース	文 数		翻訳／ ソース
	ソース	翻訳		ソース	翻訳	
秘密 (英題: Naoko)	120398	103997	0.86	10605	9293	0.88
容疑者 X の献身 (英題: The Devotion of Suspect X)	102702	89326	0.87	8864	8487	0.96
新参者 (英題: Newcomer)	94923	79411	0.84	8174	7856	0.96
ナミヤ雑貨店の奇蹟 (英題: The Miracles of the Namiya General Store)	113657	96342	0.85	10015	9785	0.98
合 計	431680	369076	0.85	37658	35421	0.94

一方作品間を比較すると、語数と文数についてソーステキスト対翻訳テキスト比の値をみたとき、語数の比には大きな差はみられないが、文数の比では『秘密』vs. *Naoko* の値が他の作品と比較して小さくなっている。(0.88 : 0.96, 0.96, 0.98)

5. データ分析の結果

5.1 品詞構成率

まずテキストの計量分析を行う際によく使用されるテキストの品詞構成率をみる。品詞の分類はソーステキストでは形態素解析エンジン Mecab の結果に沿って、そして翻訳テキストでは Stanford POS Tagger から得られた結果をもとにしたが、Stanford POS Tagger による分類の結果は語形変化も含まれるので、一般的な品詞分類、つまり名詞・動詞・形容詞・副詞・接続詞・代名詞・前置詞等にまとめて集計した。表4と表5は、それぞれの作品について10万語あたりの品詞の出現頻度と構成比率をあらわしたものである。

表4 ソーステキストの品詞構成

作 品 名	助詞	名詞	動詞	助動詞	その他	合計
秘密	32137.58 (32.1%)	30203.16 (30.2%)	16325.85 (16.3%)	13145.57 (13.2%)	8187.84 (8.2%)	100000
容疑者 X の献身	32521.28 (32.5%)	30102.63 (30.1%)	16542.03 (16.5%)	13568.38 (13.6%)	7265.68 (7.3%)	100000
新参者	31696.22 (31.7%)	30196.05 (30.2%)	16357.47 (16.4%)	14290.53 (14.3%)	7459.73 (7.5%)	100000
ナミヤ雑貨店の奇蹟	31520.28 (31.5%)	30233.07 (30.2%)	16892.93 (16.9%)	13825.81 (13.9%)	7527.91 (7.5%)	100000
平 均	31968.84	30183.73	16529.57	13707.57	7610.29	
標準偏差	390.25	48.84	225.45	415.04	347.05	
変動係数(CV)×100	1.22	0.16	1.36	3.03	4.56	

表5 翻訳テキストの品詞構成

作 品 名	動詞	名詞	代名詞	前置詞	限定詞	その他	合計
Naoko	22192.95 (22.2%)	19727.49 (19.7%)	12936.91 (12.9%)	10910.89 (10.9%)	9209.88 (9.2%)	25021.88 (25.0%)	100000
The Devotion of Suspect X	22561.18 (22.6%)	19753.49 (19.8%)	13561.56 (13.6%)	10579.23 (10.6%)	9476.52 (9.5%)	24068.02 (24.1%)	100000
Newcomer	21638.06 (21.6%)	21803.02 (21.8%)	12507.08 (12.5%)	11173.52 (11.2%)	11061.44 (11.1%)	21816.88 (21.8%)	100000
The Miracles of the Namiya General Store	22315.29 (22.3%)	19660.17 (19.7%)	13627.49 (13.6%)	10177.29 (10.2%)	9828.53 (9.8%)	24391.23 (24.4%)	100000
平 均	22176.87	20236.04	13158.26	10710.23	9894.09	23824.50	
標準偏差	338.17	905.34	462.56	372.86	708.79	1208.80	
変動係数(CV)×100	1.52	4.47	3.52	3.48	7.16	5.07	

ソーステキストの4作品の品詞構成比率に大きな差はみられない一方、翻訳テキストの4作品の中では、*Newcomer*が名詞と限定詞の構成率に関して他の3作品よりも大きくなっている。それに伴い「その他」の値は小さくなっている（表5のグレー部分を参照）。

次に各品詞の変動係数（Coefficient of Variation, 以下 CV）をみる。CVとは単位の異なるデータのばらつきや、平均値に対するデータとばらつきの関係を相対的に評価するために用いる単位を持たない値であり、標準偏差を平均値で割ることによって求められる（ここでは値を見やすくするために100倍にしてある）。翻訳テキストのCVはソーステキストのそれよりも全体的に大きかった。これは翻訳テキストの作品間のほうが品詞構成率にばらつきが出ていることをあらわしている。

次に両テキストで品詞構成率の高かった機能語のうち、高頻度の単語の使用率を調べた。機能語とは作品の内容とは密接な関係がなくテキストの中で文法的機能を持つ語をいう。今回は日本語からは品詞構成率の最も高かった助詞を選択した。英語では日本語の助詞に対応するものとして前置

表6 ソーステキストの助詞出現頻度（10万語あたり）

助詞	秘 密	容疑者 X の 献 身	新 参 者	ナミヤ雑貨 店 の 奇 蹟	平 均	標準偏差	CV×100
の	4889.62	4461.50	4349.84	4524.58	4556.38	202.31	4.44
は	4357.22	4729.26	4300.33	4174.44	4390.31	206.57	4.71
に	3627.14	3781.85	3588.17	3754.79	3687.99	82.07	2.23
を	3473.48	3839.30	3633.47	3624.59	3642.71	130.10	3.57
て	3652.05	3622.17	3541.82	3470.63	3571.67	70.91	1.99
が	2955.20	2952.26	3036.15	3053.63	2999.31	46.01	1.53
と	1932.76	1877.29	1926.82	2000.56	1934.36	43.87	2.27
し	1378.76	1396.29	1343.19	1399.69	1379.48	22.41	1.62
で	1360.49	1331.05	1337.93	1453.36	1370.70	48.95	3.57
か	1069.79	1322.29	1273.66	1138.40	1201.03	101.40	8.44

表7 翻訳テキストの前置詞出現頻度（10万語あたり）

前置 詞	Naoko	The Devotion of Suspect X	Newcomer	The Miracles of the Namiya General Store	平均	標準 偏差	CV × 100
to	3069.96	2916.27	2882.03	2881.66	2937.48	77.77	2.65
of	1582.95	1410.09	1654.41	1416.48	1515.98	105.78	6.98
in	1279.79	1053.65	1202.52	1196.80	1183.19	81.65	6.90
on	642.77	736.33	758.17	706.68	710.99	43.42	6.11
at	804.90	746.39	627.62	569.91	687.20	93.09	13.55
for	701.29	641.36	651.47	746.06	685.04	41.90	6.12
with	537.24	587.72	707.96	571.98	601.22	64.27	10.69
as	611.11	470.40	456.91	452.82	497.81	65.74	13.21
about	445.14	513.98	549.80	337.80	461.68	80.81	17.50
from	337.70	419.00	379.08	428.98	391.19	36.09	9.23

詞を選択した。両テキストで最も出現頻度の高い助詞・前置詞10語の10万語あたりの頻度を表6と表7に示す。ソーステキストでは助詞「か」のCVが唯一5以上であるのに対し、翻訳テキストでは最高頻度の前置詞“to”の

CV以外の9単語がすべて5以上であった。これは同一作者が書いたソーステキストでは多くの助詞の使用頻度が一定である一方で、翻訳テキストでは多くの前置詞の使用頻度には翻訳者間でばらつきがあることを示している。

5.2 文の長さ

文の長さの比較は計量言語学で最も使用される手法の一つであり、文の長さは作者の文章の書き方の一般的なパターンをあらわすと考えられている。本研究ではBaker (2000) や Mikhailov and Cooper (2016) で扱われた全文の長さの平均という値だけでなく、文の長さについての分布と、ソーステキストと対応する翻訳テキストの文(章)の長さの比もみる。

表3でソーステキストと翻訳テキストそれぞれの語数と文数はすでに示したが、表8はその語数を文数で割った一文の平均の長さを示したものである。

表8 ソーステキストと翻訳テキストの文の長さの平均

作 品 名	文の長さの平均	
	ソーステキスト	翻訳テキスト
秘密 (英題: Naoko)	11.35	11.19
容疑者Xの献身 (英題: The Devotion of Suspect X)	11.59	10.53
新参者 (英題: Newcomer)	11.61	10.11
ナミヤ雑貨店の奇蹟 (英題: The Miracles of the Namiya General Store)	11.35	9.85
平 均	11.48	10.42
標準偏差	0.12	0.51
CV	1.09	4.88

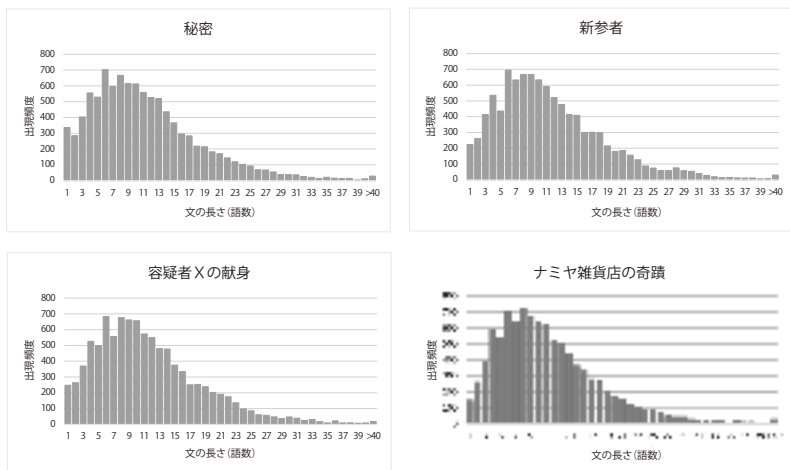


図1 ソーステキストの文の長さの頻度分布

ソーステキストと翻訳テキストの間に差はほとんどみられないが、ソーステキストは4作品の文の長さのCVは1.09であるのに対し、翻訳テキストの場合4作品の平均の長さのCVは4.88と大きくなっている。

それぞれのテキストについて文の長さをもとにその頻度の分布をあらわしたものが図1・図2である。図1が示すように、ソーステキストにおける文の長さの頻度は、4作品すべてにおいて6語から9語の間で頻度が700程度でピークとなっており、減少する割合も類似している。一方、図2が示す翻訳テキストでは、グラフの形状が4作品で大きく違っている。特に*Naoko*では4語から9語の間において、頻度550程度で水平部分が観察され、かつ他と比較して文の長さに従って頻度の減少する割合がゆるやかである。*Newcomer・The Miracles of the Namiya General Store*では6語あたりで頻度800を僅かに下回る程度でピークを迎えている点が他の2作品と異なり、かつ後者は他の3作品の分布と比較しても文の長さは全体的に短いようにみえる(表8も参照)。

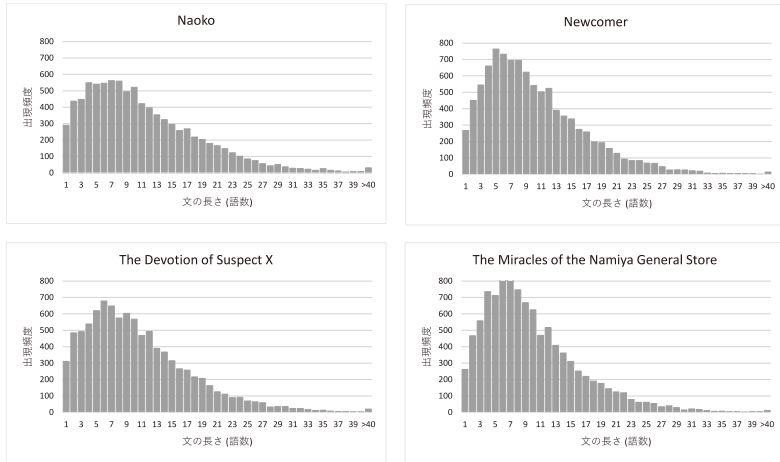
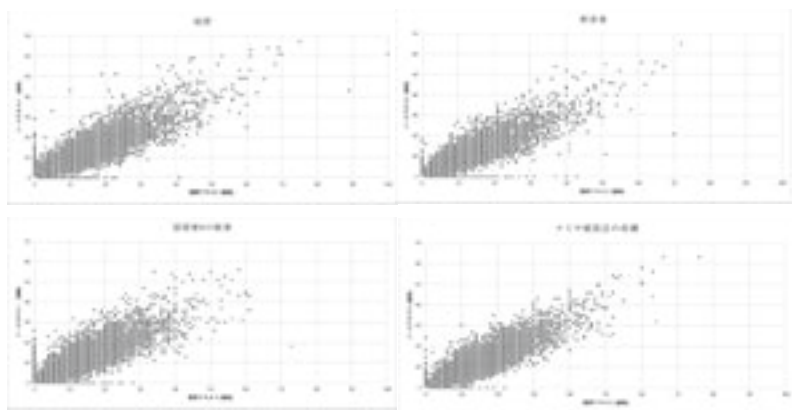


図2 翻訳テキストの文の長さの頻度分布

5.3 ソーステキストと翻訳テキストの文（章）の長さの相関

本研究で作成したコーパスは日本語テキストとそれに対応した英語テキストで構成されるパラレルコーパスであるので、それぞれ個々のソーステキストの長さに対応する翻訳テキストの長さを比較することが可能である。ソーステキストに対しての翻訳テキストの長さをあらわした散布図が図3である。ここでは文対文だけでなく意味的に対応する文章も含まれることに注意されたい（表2のNo.469を参照）。

相関係数をみると4作品とも強い正の相関がみられる ($R > 0.7$)。しかし、『ナミヤ雑貨店の奇蹟』vs. *The Miracles of the Namiya General Store* は一番相関係数が大きく、『容疑者Xの献身』vs. *The Devotion of Suspect X* が最も小さい。グラフ上からその二つを比較したとき、文章が長くなればなるほど前者の分布はまとまって細くなっている一方、後者の分布はわずかに広がっている点からも読み取ることができる。



	秘密	容疑者Xの献身	新参者	ナミヤ雑貨店の奇蹟
相関係数	0.851	0.819	0.834	0.863

図3 ソーステキストと翻訳テキストの長さの相関

5.4 文の加筆修正

翻訳テキストには、加筆された文とソーステキストに書かれていたにもかかわらず削除された文が存在する。ソーステキストから翻訳テキストに翻訳されたときに、加筆された文の数と削除された文の数を作品ごとにあらわしたものが表9である。*The Devotion of Suspect X*だけ加筆された文が多く、他の3作品は削除された文のほうが多い。

表9 翻訳テキストの加筆修正された文数

作 品 名	加筆された文の数	削除された文の数
Naoko	121	239
The Devotion of Suspect X	236	167
Newcomer	88	255
The Miracles of the Namiya General Store	72	165
平 均	129.25	206.50
標準偏差	64.11	40.90

日本小説の英訳本における翻訳者の文体的特徴

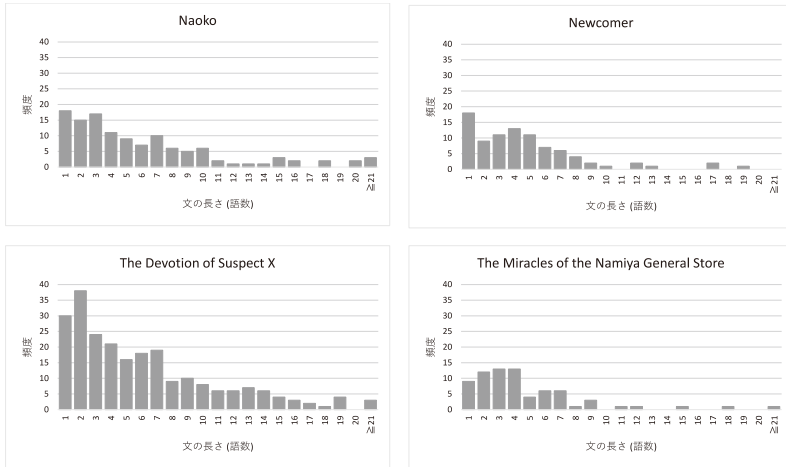


図4 翻訳テキストに加筆された英文の長さ

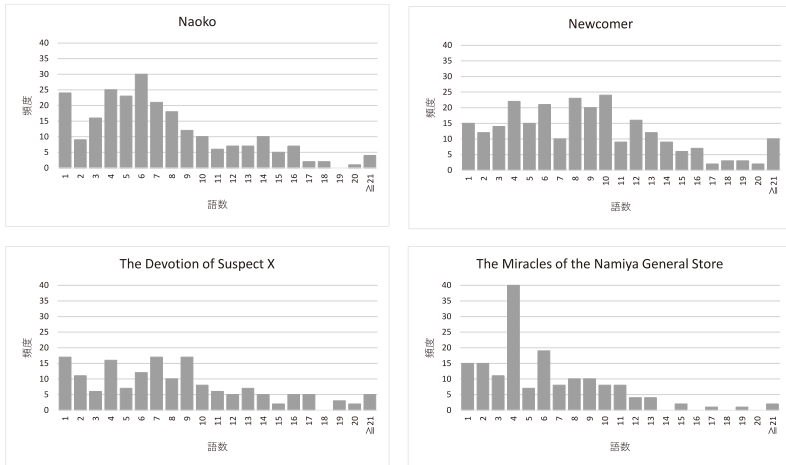


図5 翻訳テキストで削除された日本文の長さ

図4・図5は加筆された英文の長さの分布と削除されたソースの日本語の長さの分布をあらわしたものである。

図4が示すように、加筆された文の長さは全体的に短く、その多くは1語から7語程度である。*The Devotion of Suspect X*で数多く出現する長さ2語の文の加筆のほとんどは、セリフの間に登場人物の表情・動作を描写するものであった。以下に例を示す。太字が加筆された部分である。

ソーステキスト：

「脱がすのを手伝ってください」

「えっ……」

『容疑者Xの献身』 2

翻訳テキスト：

‘Help me undress him.’

Yasuko blinked. ‘What?’

The Devotion of Suspect X, 2

ソーステキスト：

「それは、あの、何となくそんな
ふうに思っただけです」

「わかりました。お時間をとらせて
て申し訳ありませんでした」

『容疑者Xの献身』13

翻訳テキスト：

‘Well, I wouldn’t say “know” so much
as that’s the impression of him.’

He nodded. ‘Very well. Sorry to
take up your time.’

The Devotion of Suspect X, 13

また *Newcomer* で加筆された1語はほぼ登場人物の台詞であった。

ソーステキスト：

「一昨日、おたくにお邪魔した
時、店の奥で寺田さんがおっしゃ
ってたじゃないですか。まさか、
あのチンピラが何かしでかしたん
じゃないだろうなって」

『新参者』第4章 3

翻訳テキスト：

‘It was something Mr Terada said in
his workroom when I was there the
day before yesterday. **Remember?**
“Bet it’s that punk kid. Probably went
and got in trouble with the law.”

Newcomer, Chapter 4:3

ソーステキスト:

「その前だと……一ヵ月以上前
になると思いますけど」

「では、十二日に買ったお客さ
んのことで、何か覚えていること
はありませんか」

『新参者』第8章 1

翻訳テキスト:

‘Before that? ... There’s a gap of a
whole month.’

‘**Good.** Now, do you recall
anything about the person who
made this purchase on the twelfth?’

Newcomer, Chapter 8:1

図5が示すように、4作品すべてにおいて、ソーステキストから1語か
ら15語程度の長さの日本語に対応する部分が、翻訳テキストでは削除され
ていた。具体例をみると連続して複数の行が削除されている場合も多い。
太字は省略された文である。

ソーステキスト:

彼は寝室に入り、本棚に載せてあ
るウォークマンを手にとった。**蓋を開
け、テープを入れる。イヤホン**を耳
に突っ込み、**テープを巻き戻す。**

これを聞くのが毎日の楽しみの
一つになっていた。『秘密』34

翻訳テキスト:

He went into his bedroom and took
his Walkman from the bookshelf.
Listening to these conversations had
become one of his daily pleasures.

Naoko, 34

ソーステキスト:

「ありえないですね。父が俺の
住所を教えるとは思えないし」

「うん。お父さんはおしえてな
いとおっしゃってる」

翻訳テキスト:

‘That seems out of the question.
There’s no way Dad would have told
her where I live. It had to be a
coincidence, pure and simple.’

「だったら、単なる偶然ってことじゃないですか」

『新参者』第5章 2

Newcomer Chapter 5:2

また *The Miracles of the Namiya General Store* において削除された文の中で、突出して多かった4語の文の多くは典型的な発言の描写であった。

ソーステキスト：

「おい、なんだよ。また返事を書く気か。からかわれているだけなんだぞ」敦也はいった。

「でも、そんなのまだわかんないし」

『ナミヤ雑貨店の奇蹟』
第1章 3

翻訳テキスト：

“Wait a sec? Are you writing back? This is just somebody screwing with us.”

“But we don’t know that for sure yet.”

The Miracles of the Namiya General Store, Chapter 1:3

ソーステキスト：

「通夜は明日。で、明後日がお葬式。お兄ちゃん、いつ帰ってこれる？」栄美子が訊いた。

克郎は受話器を持った手でカウンターに肘をつき、もう一方の手で頭を掻いた。

『ナミヤ雑貨店の奇蹟』
第2章 1

翻訳テキスト：

“The tsuya ceremony’s tomorrow, and the funeral’s the day after. How soon can you be here?”

Katsuro propped his elbow on the counter and scratched his head with his free hand.

The Miracles of the Namiya General Store, Chapter 2:1

5.5 語彙の豊富さ

5.5.1 STTR

テキストで使用される語彙の豊富さもテキストの文体の特徴をあらわすと考えられている。語彙の豊富さの計量的な指標として最もよく知られているのはタイプ・トークン比 (type/token ratio) である。これはテキスト中にあらわれる異なる単語の数 (異なり語数・タイプ数) を出現する総単語数 (延べ語数・トークン数) で割り、パーセンテージであらわされる。しかしテキストの一般的な特徴として、テキストが長くなればなるほど、一度しか出現しない単語が多くなるという事実から、このタイプ・トークン比はテキストの長さに強く依存することがわかっている。

Baker (2000) や Mikhailov and Cooper (2016) ではこのタイプ・トークン比の欠点を修正した STTR (standardized type/token ratio・標準化タイプ・トークン比) を使用している。これはテキストを任意の語数 (彼らの使用したソフトウェア WordSmith Tools での初期設定値は1000語) に分割し、その分割したサブテキストでタイプ・トークン比を求め、その平均値を求める方法である。

ソーステキストと翻訳テキストそれぞれの作品について1000語で分割したサブテキストを作成したのち STTR を求め、まとめたものが表10である。ここでは形態素解析エンジン Mecab と Stanford POS Tagger から得られた

表10 ソーステキストと翻訳テキストの STTR

作 品 名	STTR	作 品 名	STTR
秘密	34.18	Naoko	34.81
容疑者 X の献身	32.63	The Devotion of Suspect X	33.77
新参者	33.20	Newcomer	35.42
ナミヤ雑貨店の奇蹟	34.18	The Miracles of the Namiya General Store	34.25
平 均	33.55	平 均	34.56
標準偏差	0.66	標準偏差	0.62
CV × 100	1.98	CV × 100	1.78

辞書見出し語（原形・Lemma）を使用して集計した。

ソーステキストと翻訳テキストでは言語が日本語と英語で文法構造も大きく異なるので、それらを直接比較することはできない（Mikhailov and Cooper 2016, p. 118を参照）。一方それぞれテキストで4つの作品間のばらつきのほうをみると、CVはごくわずかに翻訳テキストのほうが小さかった（1.98 : 1.78）。

5.5.2 YULE の K 特性値

TTRもしくはSTTRはテキストの異なり語数と延べ語数から語彙の豊かさを求めるが、それ以外に出現する単語すべての頻度にもとづき語彙の豊かさを計算する方法もある。その中で本研究ではYule (1944) で提案された K 特性値（Characteristic K ）を使用する。この値は、 N をトークン数、 n を頻度、 $V(n, N)$ をトークン数 N のテキストにおいて頻度 n の語の異なり語数としたとき、次の式から求められる。この値は小さいほど、テキストの語彙が豊富であることをあらわす。

$$K = 10^4 \frac{\left[\sum_{n=0}^{all} n^2 V(n, N) \right] - N}{N^2}$$

表11はソーステキストと翻訳テキストのそれぞれの作品の K 特性値である。表10での結果と同様、CVの値がソーステキストの作品間よりも翻訳テ

表11 ソーステキストと翻訳テキストの K 特性値

作 品 名	K	作 品 名	K
秘密	167.24	Naoko	118.45
容疑者Xの献身	168.87	The Devotion of Suspect X	118.76
新参者	160.47	Newcomer	114.40
ナミヤ雑貨店の奇蹟	162.55	The Miracles of the Namiya General Store	117.57
平 均	164.78	平 均	117.30
標準偏差	3.40	標準偏差	1.73
CV × 100	2.07	CV × 100	1.47

キストの作品間のほうが小さくなっている (2.07 : 1.47)。

5.5.3 語彙密度

語彙密度 (Lexical Density) は、STTRや*K*特性値のような語彙の多様性という観点から語彙の豊富さを測るものとは異なり、テキストの中の内容語の割合を計算することで、そのテキストにおける情報の密度をあらわす。Halliday (1985) はテキストに含まれる内容語の総数をテキストに含まれる節の総数で割ることによって求めると定義したが、本研究ではUre (1971) の定義に従って、テキストに含まれる語の総数 (延べ語数) で割ることによって求めることとした。本研究では、一般的に用いられている定義と同様に内容語を名詞・動詞・形容詞・副詞とした。表12はその結果をあらわしたものである。語彙密度については、CVは翻訳テキストの作品間よりもソーステキストの作品間のほうが小さいが、その差はわずかであった (0.74 : 1.11)。

表12 ソーステキストと翻訳テキストの語彙密度

作 品 名	LD	作 品 名	LD
秘密	51.47	Naoko	55.60
容疑者 X の献身	50.93	The Devotion of Suspect X	55.05
新参者	50.78	Newcomer	55.91
ナミヤ雑貨店の奇蹟	51.71	The Miracles of the Namiya General Store	54.29
平 均	51.22	平 均	55.21
標準偏差	0.38	標準偏差	0.62
CV × 100	0.74	CV × 100	1.11

5.6 読点とカンマ

金 (2021) によると日本文による読点は、並立する語句の間に打つ読点以外、明確な基準がないため書き手の特徴が出やすいという。また Baker (1996) によれば、翻訳作業中、翻訳者は英文をよりわかりやすくするため、より一般的なカンマの打ち方に変更するという。これらの特徴を検証する

ため、まず基礎統計量としてソーステキストと翻訳テキストの各作品の読点・カンマの数を調べた。記号を除いた単語100語あたりに出現する頻度を表13で示す。ソーステキストと翻訳テキストのCVを比較すると、翻訳テキストの値がより大きいことがわかる（6.13：10.08）。

次にソーステキストと翻訳テキストにおいて、最初の読点およびカンマが文の先頭から何番目に出現するかをあらわしたものが図6・図7である。ここではすべての記号を含めたコーパスを使用した。図6のソーステキストでは、それぞれ頻度の違いによって形状が異なっているが、3語目の位置にあらわれることが最も多く、以降5語目をピークにして減少している点は共通である。図7の翻訳テキストの作品では、3語目の位置にあらわれることは共通だが、*Naoko*と*The Devotion of Suspect X*が極端に多く（頻度700～800）、他の2作品はそれほど多くない。また2番目のピークも*Naoko*と*Newcomer*は7語目、残りの2作品は5語目と異なっている。しかしグラフの形状からだけではソーステキストの作品間と翻訳テキストの作品間との間に大きな差異がみられると断言することは難しいように思われる。

図6・図7で最大頻度である読点・カンマが3語目にあらわれるものの

表13 ソーステキストの読点と翻訳テキストのカンマの出現数・頻度

作 品 名	読点	出現頻度 (単語100語 あたり)	作 品 名	カンマ	出現頻度 (単語100語 あたり)
秘密	5559	4.70	Naoko	5918	5.69
容疑者Xの献身	5004	4.87	The Devotion of Suspect X	4823	5.40
新参者	4922	5.19	Newcomer	3504	4.41
ナミヤ雑貨店の 奇蹟	6253	5.50	The Miracles of the Namiya General Store	4566	4.74
平 均	5459.5	5.06	平 均	4702.75	5.06
標準偏差	539.87	0.31	標準偏差	858.34	0.51
CV×100	9.39	6.13	CV×100	18.25	10.08

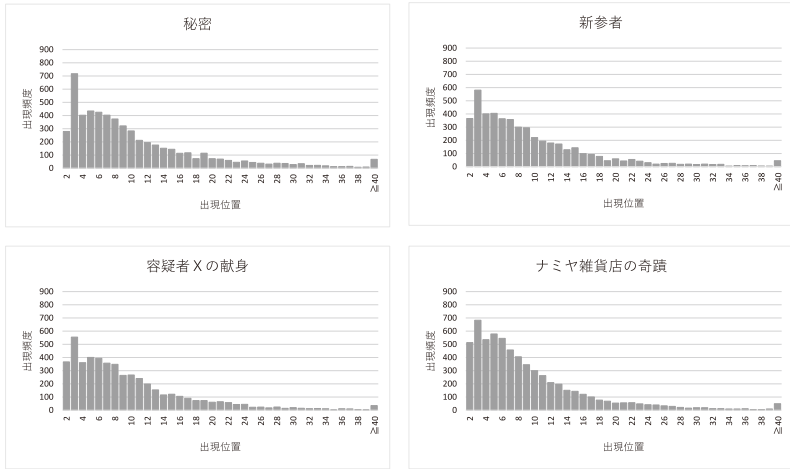


図6 ソーステキストの読点の出現位置とその頻度

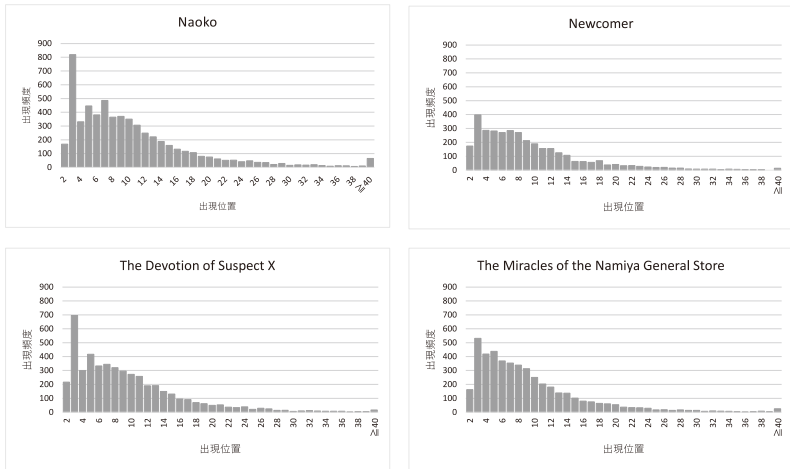


図7 翻訳テキストのカンマの出現位置とその頻度

中から、頻度が10以上の文頭から3語目までの語列をリストにしたものが表14である。記号も含めて検索したので、それぞれの作品において記号を

表14 文頭から3語目が読点・カンマとなる語列の頻度

秘密		容疑者Xの献身		新参者		ナミヤ雑貨店の奇蹟	
文頭の表現	頻度	文頭の表現	頻度	文頭の表現	頻度	文頭の表現	頻度
「いや、	39	「いや、	36	「じゃあ、	27	「じゃあ、	25
「あ、	35	「じゃあ、	25	「いえ、	24	「いや、	24
「あつ、	34	「いえ、	25	「いや、	22	「まあ、	20
「えつ、	30	「あ、	24	「あ、	16	「うん、	13
「ええ、	26	「まあ、	21	「あつ、	13	浩介は、	11
「ああ、	25	「ええ、	16	加賀は、	11	「何だ、	10
「いえ、	22	「うん、	13	「まあ、	10	「だから、	10
「うん、	21	「あの、	11				
「じゃあ、	16	「はい、	10				
平介は、	14						
「ええと、	14						
「はい、	13						
「でも、	11						

Naoko		The Devotion of Suspect X		Newcomer		The Miracles of the Namiya General Store	
文頭の表現	頻度	文頭の表現	頻度	文頭の表現	頻度	文頭の表現	頻度
“ Oh,	107	‘ Well,	98	‘ Oh,	49	“ Yeah,	37
“ No,	84	‘ No,	78	‘ Yes,	38	“ No,	25
“ Well,	64	‘ Oh,	38	‘ No,	33	“ Hey,	25
“ Yeah,	63	‘ Yes,	31	‘ Well,	18	“ Well,	21
“ Yes,	59	Of course,	22	I mean,	12	I mean,	16
“ Okay,	31	‘ Yeah,	18	‘ Yeah,	12	“ Oh,	11
“ Ah,	15	‘ Right,	15	Sure enough,	11	“ Wait,	10
“ Uh,	13	‘ So,	13	‘ Anyway,	10		
“ Hey,	13	‘ Actually,	12				
“ Dad,	11	In fact,	10				
In fact,	10	After all,	10				
“ So,	10	‘ Hey,	10				

含む「記号+間投詞+読点・カンマ」という登場人物の発話の先頭をあらわす語列が上位に多くあらわれた。表14にあらわれる各作品の語列の数はソーステキストと翻訳テキストの間でおおよその比例関係がみられるが、頻度の数は大きく違っており、翻訳テキストのほうが間投詞を含んだ語列の頻度が高い。また日本語の間投詞と英語の間投詞には対の関係はみえない。つまりソーステキストにあらわれる間投詞は、翻訳テキストで書かれる言語（英語）での文脈によって翻訳者が訳語を変化させているようである。

6. 考 察

以上のように、「品詞構成率と高頻度機能語使用率」、「文の長さとその頻度分布」、「ソーステキストと翻訳テキスト間での対応する文（章）の長さの相関関係」、「翻訳テキストでの文の加筆修正」、「語彙の豊富さ」、「読点とカンマの頻度と出現位置」の6つの計量言語学的手法を用いて、翻訳テキストにおける文体の特徴の特定を試みた。

ソーステキストと翻訳テキストそれぞれで各作品間の値のばらつきをあらわす次の7つの指標——「品詞構成率と高頻度機能語使用率」、「文の長さの頻度分布」、「3種類の語彙の豊富さの指標（STTR・*K*特性値・語彙密度）」、「読点とカンマの頻度」——では、STTRと*K*特性値という語彙の豊富さをあらわす指標を除いたすべての指標で、ソーステキストと比較して翻訳テキストのほうが各作品間にばらつきがあった。これは同一の作家によって執筆されたソーステキストより、それが翻訳されるときに、異なる翻訳者の文体が意識的もしくは無意識的に翻訳テキストに付加されることをあらわしている。

金（2021）によれば、書き手があまり意識しない助詞の使用パターンは書き手の特徴が明確にあらわれるとされている。類似した機能語である前置詞の選択のパターンも翻訳者それぞれの文体と考えるのが妥当であろう。

同様に書き手があまり意識しない（もしくはできない）品詞の構成率や語彙密度も翻訳者自身に起因する文体の特徴をあらわすと考えられる。

2.2節で指摘したように Baker (1996) では、「簡素化」と呼ばれる翻訳行為の普遍的特徴によって、翻訳された文の長さは分割され短くなる傾向があるとされた。図 1 と図 2 をみると、他の作品と比較したとき、Sam Bett による *The Miracles of the Namiya General Store* はその特徴が観察される。

日本語の読点と英語のカンマは、語句の切れや続きを明確にするために、文中の意味の切れ目に付与される記号という点では類似しているが、日本語と英語では文法が大きく異なるため、その相関関係は低いと思われる。またカンマの打ち方が、前述した金 (2021) の指摘にあるような日本語の読点のように、書き手の特徴があらわれやすいものだとすると、本研究で示したカンマの頻度も翻訳者の文体の一つの特徴量となるかもしれない。さらに興味深いのは、表 14 で示した読点・カンマの位置から得られた会話の先頭に書かれる間投詞についてである。翻訳テキストの間投詞はソーステキストのそれよりも頻度が多く、各作品で選択のパターンにばらつきがある。これは Baker (1996) で指摘されたソーステキストにみられる表現を翻訳される言語に合わせて変化させるという「正規化」が、ここでの翻訳テキストでも観察されたと考えられる。間投詞をどの程度文頭に置くか、その時の間投詞は何にすべきかについて、英語の言語習慣とその文化に合わせて翻訳されたと考えるのが妥当であろう。

本研究では「語彙の豊富さ」の指標 STTR と K 特性値については、ソーステキストよりも翻訳テキストのほうが作品間のばらつきが小さかった。これは、Baker (1996) での翻訳されたテキストが極端な計量言語学的特徴を失う傾向があることをさす「平準化」の特徴を示しているのではないかと考えられる。しかし Mikhailov and Cooper (2016) では、STTR を用いた語彙の豊富さの調査では、翻訳テキストにおいて翻訳者自身の文体が反映

されると報告しており、本研究の結果とは異なる。この違いは調査した翻訳者の傾向の違いとも、ソーステキスト・翻訳テキストの言語の違いとも考えられるが、紙面の都合上、その特定は今後の課題としたい。

「翻訳テキストでの文の加筆修正」という行為ひとつひとつは、翻訳者の意識的作業であると考えられるが、翻訳者はそのパターンには無意識かもしれない。Alexander O. Smith による *The Devotion of the Suspect X* の具体例でみた加筆は会話の中で挿入され、発話者が誰なのかや発話者の動作を描写していた。Giles Murray 翻訳の *Newcomer* では登場人物の発話の中に間投詞が一語加筆された具体例をみた。これは翻訳テキストが書かれる言語の文章や発話のリズムに合わせて変更を加えた「正規化」と考えられる。一方でいくつかの翻訳テキストで具体例としてみられた削除は、ソーステキストが持つ冗長さを解消して訳す「簡素化」と考えることができる。

「ソーステキストと翻訳テキスト間での対応する文(章)の長さの相関関係」でみられた『容疑者 X の献身』vs. *The Devotion of the Suspect X* の相関係数が小さかったのは、Baker (1996) の指摘にあった、翻訳においてソーステキストに暗示される情報を翻訳テキストに表出する「明示化」の目的によって翻訳者の加筆が多かったためとも考えられる。一方で『ナミヤ雑貨店の奇蹟』vs. *The Miracles of the Namiya General Store* の相関関係が高かったのは、その翻訳テキストでの文の加筆や削除もそれほど多くはなかったという事実と併せて、ソーステキストを忠実に翻訳していることを示しているといえるだろう。これらも翻訳者の文体の違いと捉えることもできる。

7. ま と め

本稿は日本小説の英訳テキストでの翻訳者のスタイルの特徴を特定するにあたって、ソーステキストのバイアスをできる限りなくすため、東野圭吾作の4小説とその英訳本からパラレルコーパスを構築し計量言語学的方

法を用いて分析を行った。用いた手法は「品詞構成率」,「高頻度機能語使用率」,「文の長さとその頻度分布」,「ソーステキストと翻訳テキスト間での対応する文(章)の長さの相関関係」,「翻訳テキストでの加筆修正」,「語彙の豊富さ(STTR・K特性値・語彙密度)」,「読点とカンマの頻度と出現位置」の7つであり,得られた結果それぞれに対して,翻訳者自身の文体から生じたものか,Baker(1996)にある翻訳行為にみられる普遍的特徴から生じたものかを考察した。

最後に6節で考察した結果をまとめておく。

翻訳者自身の文体によって差が生じると考えられるもの:

「品詞構成率」「高頻度機能語使用率」「語彙密度」「カンマの数」のばらつき

普遍的特徴によって文体に差が生じると考えられるもの:

明示化と考えられるもの:

「ソーステキストと翻訳テキスト間での対応する文(章)の長さの相関関係」のばらつき

簡素化と考えられるもの:

「翻訳テキストでの文の削除」「短文化」

正規化と考えられるもの:

「翻訳テキストでの文の加筆」「問投詞の加筆」

平準化と考えられるもの:

「語彙の豊富さ(STTR・K特性値)」の中庸化

残念ながら,計量言語学的方法論だけでは調査で見つかった差異の意味解釈を行うことは難しい。本研究の考察の妥当性を裏付けるためには,さらなる言語データからの検証や異なる観点からの研究が必要となるだろう。

注

- 1) 著作権の問題については西村（2002）を参照。本の購入者が個人として使用する場合は著作権には抵触しないと判断されている。
- 2) 実際に入力したデータは以下のとおりである。

『秘密』：文春文庫 2010年7月5日46刷 Naoko: Vertical Inc. 2004 First American edition

『容疑者Xの献身』：文春文庫 2008年8月15日2刷 *The Devotion of Suspect X*: Abcus 2012 Paperback edition

『新参者』：講談社文庫 2013年8月9日1刷 *Newcomer*: Abcus 2019 Paperback edition

『ナミヤ雑貨店の奇蹟』：角川文庫 2017年8月10日32刷 *The Miracles of the Namiya General Store*: Yen Press June 2021 First paperback edition
- 3) 『現代日本語書き言葉均衡コーパス』利用の手引 第1.1版 第5章 形態論情報 (https://ccd.ninjal.ac.jp/bccwj/doc/manual/BCCWJ_Manual_05.pdf) を参照。

参考文献

- Baker, Mona. 1996. Corpus-based translation studies: the challenges that lie ahead. In Somers, H. (ed.) *Terminology, LSP and Translation: Studies in Language Engineering*, in Honour of Juan C. Sager. Amsterdam: John Benjamins, 175-186.
- . 2000. Towards a methodology for investigating the style of a literary translator. *Target*. 12: 2, 241-266.
- Halliday, Michael. 1985. *Spoken and Written Language*. Deakin University.
- 金明哲 2021. 「スタイロメトリー分析」金明哲・中村靖子（編）『文学と言語コーパスのマイニング』1-22. 東京：岩波書店.
- Mikhailov, Mikhail. & Cooper, Robert. 2016. *Corpus linguistics for translation and contrastive studies: A guide for research*. Routledge Corpus Linguistics Guides. London and New York: Routledge.
- 西村公正 2002. 「「関西外大コーパス B—日英パラレルコーパス」の概要—計画から作成作業、およびその内容—」『英語コーパス研究』第9号, 37-43.
- Olohan, Maeve. 2004. *Introducing Corpora in Translation Studies*. London and New York: Routledge.
- Ure, Jean. 1971. Lexical density and register differentiation. In G. Perren and J.L.M. Trim (eds), *Applications of Linguistics*, London: Cambridge University Press. 443-452.

Yule, George. Udny. 1944. *The statistical study of literary vocabulary*. Cambridge: Cambridge Univ. Press.