様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２８ 年　 ９ 月 １１ 日現在

機関番号：６２６１８
研究種目：若手研究(B)
研究期間：2013～2015
課題番号：２５７７０１７８

研究課題名（和文）近代口語文翻訳小説コーパスの構築と計量的文体研究

研究課題名（英文）Compilation of a Corpus of Translated Works of the Mid-Meiji Era and Its Quantitative Analysis

研究代表者
小西　光（Konishi, Hikari）

大学共同利用機関法人人間文化研究機構国立国語研究所・コーパス開発センター・プロジェクト非常勤研究員

研究者番号：３０６４６５９２

交付決定額（研究期間全体）：（直接経費）　　3,100,000円

研究成果の概要（和文）：本研究では明治20年代に発表された言文一致体翻訳小説（６作品）のコーパス構築を行った。本コーパスには、形態素情報・節情報が付与されている。
いわゆる「言文一致体」を生み、近代口語体の確立にいたる過程で重要な役割を果たしたと考えられる明治20年代の翻訳小説に着目し、形態素情報をもとに統計的な分析を行った。比較には明治末期から大正期に創作された長編小説と短編小説を用いた。その結果、形態素情報レベルのn-gramは近似の値を示すという結果が得られた。

研究成果の概要（英文）：We compiled a corpus of six translated works of the mid-Meiji era and one of each from the late Meiji and Taisyo eras.
This research provides an overview of the corpus and the findings of the statistic survey conducted, such as the length of sentence, the similarity between the works, and so on. The survey shows that there are no significant differences between the two sets of works in the number of bunsetsus in a sentence, the rate of POS tagging, Modifier Verb Ratio (MVR), or in their cosine similarities. The quantitative results indicate the similarities between the translated works of the mid-Meiji era and those of the late Meiji and Taisyo eras. It also supports the assumption that the colloquial style had already become familiar through translation works and read by people of the mid-Meiji era. However, as an exception, statistics show that the variance of the conjunctive postposition in a sentence tends to decrease with the passage of time.

研究分野：日本語学

キーワード：文体　近代口語体　言文一致体　翻訳小説　コーパス

１．研究開始当初の背景

日本語の文体模索期ともいえる明治前・中期の多様な文体について、これまでは「代表例」とされる言語資料やその特徴が挙げられるものの、定量的分析による具体的かつ有効な指標は研究が進められている段階にある。日本語文体の大きな転換点と位置づけられる「言文一致運動」で生まれた言文一致体の研究では、文末辞「ダ」「デアル」「デス・マス」といった文末に限定した機能的類型で論じられてきた。文末辞は文体論において重要な特徴量ではあるが、文末辞のみに着目しただけでは、言文一致体の全体像を的確にとらえることは出来ず、論じるうえで不十分である。文末辞以外に文法・語彙・文構造など様々なレベルで文体をつかさどる特徴を体系的に調査することが重要である。

統計的手法による文体類型化の指標については、近現代文学を広範囲に扱ったものがある。近代語に限ると、近年雑誌・新聞・教科書・翻訳小説等多様なメディアにおける近代文語文分類の有効な指標の指摘や類型化研究が進められつつあるが、近代口語文については文語文同様ジャンル横断的調査が必要である。

これを可能にするのは、当時の文章表現を網羅的に収集した『近代語のコーパス』となる。国立国語研究所では『太陽コーパス』『近代女性雑誌コーパス』『明六雑誌コーパス』など、文語体・口語体混在の近代語のコーパスが整備されている。しかしながら、現代語では均衡コーパスが整備されている一方、近代語では雑誌メディアを中心としたものであり、近代文体調査を行うには偏りがある。

そこで、本研究では以下のような目的を設定した。

２．研究の目的

本研究の主軸は以下の２点とした。

(1) 基礎的研究資料として明治 20 年代に発表された口語文（言文一致体）による翻訳小説の形態論情報付きコーパスを構築する。

(2) 言文一致体翻訳小説と近代口語文体の文体的距離を定量的分析により明らかにする。

１点目の『近代口語文翻訳小説コーパス』を構築するにあたり、対象資料は「言文一致体翻訳（直訳）小説」とした。本研究は近代口語文体成立以前の調査資料を言文一致体翻訳小説に限った点に特色がある。その理由を次に述べる。言文一致運動を担った文学者は外国語に通じていたという背景があり、「言文一致体」の模索も海外文学作品の翻訳を通じてであったことが明らかにされている。しかし、その影響力・時代的な評価に反して翻訳小説は言語資料として二流とされ、近代文体研究において調査・分析は充分になされてこなかった。言文一致体は翻訳での試行から創作の試行への流れがあり、実際の言文一致体翻訳（直訳）小説にあたると、一文の構造において近代口語文との近似性を見て取れる。ここに問題点を見出し、本研究は近代口語文体確立以前の言文一致体翻訳小説に着目し、構文レベルでの研究に耐えうるコーパス構築を行う。

２点目の「言文一致体翻訳小説と近代口語文体の文体的距離を定量的分析により明らかにする」という国語史研究について述べる。明治 20 年代に発表された言文一致体翻訳小説は、その翻訳行為を通していわゆる「言文一致体」を生み、近代口語体の確立にいたる過程で重要な役割を果たした言語資料である。通時的な文体研究において大きな転換期の言語資料にもかかわらず二流資料とされ、計量的な分析は不十分である。本研究では基礎的研究資料として言文一致体翻訳小説の形態論情報付きコーパスを構築し、統計的手法に基づく文体分析を行う。

３．研究の方法

本研究の核をなすコーパスの構築において、調査対象資料は二葉亭四迷訳「あひゞき」「めぐりあひ」、森鴎外訳「洪水」「緑葉歎」「玉を懐いて罪あり」、内田魯庵訳『小説 罪と罰 巻一』の６作品を選定した。選定条件は(1)明治 20 年代発表 (2)言文一致ダ・デアル体作品 (3)翻案ではなく直訳で翻訳された作品 (4)翻訳原文不問の４点である。この６作品に加えて、明治後期と大正期に発表された小説２作品島崎藤村『破戒』と森鴎外「高瀬舟」を比較資料としてコーパスに加えた。

データは形態素解析し人手による修正を行った。その際に重要となる単位の認定や品詞体系は国立国語研究所開発の形態素解析辞書「UniDic」によることとした。また形態素情報の整備後に文節境界・節情報を付与してコーパス構築を行った。

文体分析では、近年研究が進められている現代日本語の文体研究などを利用しながら節単位に着目し、比較資料の２作品や雑誌を中心とする近代語コーパス、『日本語書き言葉均衡コーパス』と比較しつつ、近代口語文体確立以前の構文構造を計量的に捉える。

分析の対象は、文単位から、節、短単位までを範疇とする。これは、従来の言文一致体で論じられた文末表現を扱うためであり、また明確な論理構造を意識した近代的文章への発展を明示する指標としての妥当性・有用性を明らかにするためであり、これまで短単位ベースで進められている現代語研究との比較を行うためである。また本研究で「翻訳というリミテーション」の近代日本語への影響を定量的に把握するため、言文一致体翻訳小説の言語資料としての価値を評価していく。

４．研究成果

本研究の目的２点より、(1)コーパス構築の概要 (2)翻訳小説と近代口語体創作小説の文体分析の結果を以下に示す。

(1) 「近代口語文翻訳小説コーパス」の概要

「近代口語文翻訳小説コーパス」は、主に「近代口語 UniDic」で形態素解析をし、結果を人手により読み・品詞等の修正し形態論情報「短単位」を整備した。構築されたコーパスは、翻訳小説 61,223 短単位、創作小説 27,896 短単位の計 89,119 短単位から構成されている。翻訳小説の一文あたりの文節の数は、平均 12.57 文節、創作小説は 11.39 文節となっており、文の長さに極端な差は見られない。品詞比率においても、全作品近似の比率を示している。これら基礎統計量からは、明確な差異は指摘できない。このことから、基礎統計量という指標では選定資料の８作品は類似した基礎統計量を有する資料ということが明らかとなった。

(2) 翻訳小説と近代口語体創作小説の文体分析の結果

先行して構築が完了した内田魯庵訳『小説 罪と罰 巻一』(以降、『罪と罰』とする）は、文・連用節を単位として分析を行った。分析は、次の４点で行った。

①文末の品詞性
②文末の時制
③文末辞の使用状況
④従属節の分布

結果、『罪と罰』は全体の 8.5 割を動詞述語文が占め、名詞述語文が１割を占める。そして名詞述語文のうち約八割が文末辞（主に「である」）を用いている「である体」翻訳小説である。一方、それに伴い体言止めやデ止め等、文が不完全な形態で終了する比率は減少している。

また従属節の定量的な分布調査では、『罪と罰』は前近代的な文章要素とされる戯作脈的な要素や一文中で同一の従属節を複数回用いる累加的な表現の２点が減少していることを確認した。具体的には、条件節トや理由節カラ・ノデが二葉亭四迷『浮雲』や「あひゞき」初訳ではふたつの文を連接するために用いられているのに対し、『罪と罰』では本来の原因や条件・仮定を示す用法しか確認されない。また欧文脈の特質とされる累加的表現についても、『罪と罰』では文中に並列節ガや並列節デを挿入することで一文の流れの中に区切りを作り、単一な形式の連続による単調な文の展開を回避している。

『近代口語文翻訳小説コーパス』全体では、基礎統計から算出された文の長さや文節の長さ、および品詞 比率、また品詞比率を用いた「相の類（形容詞・形状詞・副詞・連体詞）の語数÷用の類（動詞）の語数×100」で表される MVR (Modifier Verb Ratio)（図１）、文書間類似度、そして接続助詞の配列である。比較に BCCWJ 図書館サブコーパス（LB）にてサンプリングされた一部「小説」のデータを用いた。

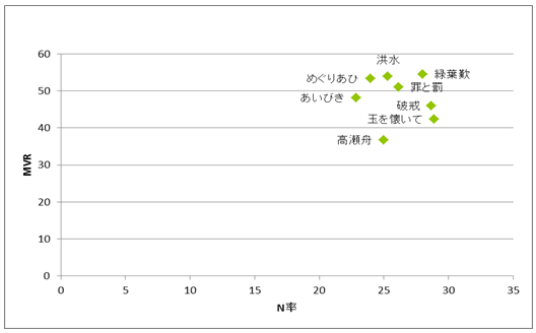

図 1　名詞率（N 率）に対する MVR

基礎統計量による文の長さや文節の長さ、品詞比率、そして品詞比率を用いた MVR については、有意な差がないという結果を得た。これは、翻訳小説６作品と近代口語体創作小説２作品の文書が、計量可能な観点では類似していることを意味している。収録された翻訳小説６作品は、筆者が純粋な読者として通読しても少々読みにくいと感じる程度の文章で、本研究の調査結果によると、当時活躍していた他の作家たちのそれとは比べものにならないほど“現代的”であったと言うことができる。

文書間類似度についても、品詞・語彙素・書字形、それぞれにユニグラム・バイグラムと粒度をこまかくしていったが、いずれも相互に文書間類似度が高く（一番、粒度のこまかい書字形分布においても低くはない)、多くの品詞もしくは語を共有していることになる。こちらも基礎統計同様、あまり差のない結果となった。

表 1 一文中の接続助詞数と接続助詞テの相関

データの個数 / 接続助詞タイプ数	1 文中に含む接続助詞の数																	
1 文中のテの数	1	2	3	4	5	6	7	8	9	10	11	12	13	14	21	24	0	総計
0																		
1		1	28	76	25	2	1											133
2			1	31	73	15	3											123
3				1	27	33	13	3										77
4					1	27	19	5										52
5						1	13	9	5		1							29
6							1	10	5	2		1						19
7								1	4	5	3	2	1					16
8									1	2		1	1					5
9										1								1
10											1			2				3
11													1					1
13														1				1
15															1			1
17																1		1
総計	1	29	108	126	78	50	28	15	10	5	4	2	1	3	1	1		462

しかし、接続助詞の配列においては、作品ごとに差異を見出すことができた。一文中に含む接続助詞の数について、その配列タイプで頻度を出し、分布を調べた。すると、四迷訳「あいびき」「めぐりあひ」や鴎外訳「玉を懐いて罪あり」「洪水」については、一文の中で接続助詞が８個や９個以上出現する文があり、文の語数にばらつきが大きいという結果になった。また、文の語数（接続助詞の数）と接続助詞「て」の数には正の相関があることも明らかとなった。（表１）

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　2　件）
①小西　光、近代口語文翻訳小説コーパス構築の概要と計量的分析、国立国語研究所論集、査読有、Vol.11、2016（刊行予定）

②小西　光、内田魯庵訳・翻訳小説『罪と罰』の言文一致文体：「文」を指標とした文体分析、上智大学国文学論集、Vol.47、2014、pp.19-38、http://repository.cc.sophia.ac.jp/dspace/bitstream/123456789/35861/2/200000020494_000133000_19.pdf

〔学会発表〕（計　2　件）
①小西　光、近代翻訳小説を資料とした品詞比率と文書間類似度による明治中期口語文体分析、第７回コーパス日本語学ワークショップ、2015、国立国語研究所
http://www.ninjal.ac.jp/event/specialists/project-meeting/files/JCLWorkshop_no7_papers/JCLWorkshop_No.7_31.pdf

②小西　光、内田魯庵訳・翻訳小説『罪と罰』における従属節から見た近代口語文体、日本語学会 2013 年度秋季大会、2013、静岡大学

〔図書〕（計　0　件）

〔産業財産権〕
○出願状況（計　0　件）

名称：
発明者：
権利者：
種類：
番号：
出願年月日：
国内外の別：

○取得状況（計　0　件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等

６．研究組織
(1)研究代表者
小西　光（KONISHI、 Hikari）
人間文化研究機構・国立国語研究所・プロジェクト非常勤研究員
研究者番号：　30646592

(2)研究分担者
(　　)

研究者番号：

(3)連携研究者
(　　)

研究者番号：