

<研究ノート>

## n-gram の分布から見た翻訳者の言語使用の特徴

### Exploring a translator's linguistic features through n-gram distributions

鄧 敏君

(台湾致理技術学院 応用日本語学科)

#### Abstract

This paper will delve into translated works by Musa Liu, who has been translating Japanese works of literature into Chinese for nearly 50 years and has published over 50 translated works in Taiwan. Focusing on n-grams which are contiguous sequences of *n* words from Japanese source texts and Chinese translated texts, this paper will examine Musa Liu looking at her linguistic patterns to determine whether these patterns exist across various translations from different source texts. Through conducting a quantitative and qualitative analysis, this study endeavors to pinpoint and discuss Liu's individual linguistic features while providing insight on the "voice" or "marks" of a translator and the nature of translation.

#### 1. はじめに

コーパス言語学の研究手法を用いた翻訳テキストの特徴を掘り下げる研究(いわば corpus-based translation studies、略して CTS)は、Baker(1993)によって提唱されて以来、大規模なパラレルコーパスや翻訳コーパスが構築され、翻訳テキストが持つ普遍的特徴(translation universals)、あるいは翻訳行為の規則性や規範について数多くの研究が行われてきた。近年では、翻訳者個人のスタイルや言語使用の傾向・パターンへの関心も増してきている。たとえば、Baker(2000)、Munday(2008)、Saldanha(2011a, 2011b)などがある。というのも、翻訳者が翻訳活動において大役を果たしているゆえ、翻訳者個人のスタイルや言語使用の特徴を究明しなければ、翻訳テキストの一般的な特徴や翻訳行為の本質を明らかにすることが難しいからであろう。本研究ノートは台湾の日本文学翻訳者である劉慕沙(LIU Musa, 1935-)の翻訳作品の文体に関する研究の一部として、翻訳テキストにおける言語使用の特徴を n-gram という計量的な手法で分析したものである。1960年代から日本文学を台湾に紹介してきた劉は、いままで50作近くの翻訳作品を出版しており、台湾の代表的な日本文学の翻訳者の一人と言っても過言ではない。本研究ノートは、劉慕沙の翻訳作品における n-gram の分布を報告し、異なる原文作品に対して劉がどのように一致した「痕跡」を残したかに注目する。

#### 2. 先行研究:n-gram の手法と翻訳文体の研究

n-gramとは隣接しているn文字あるいはn語の文字列を指す<sup>1)</sup>。たとえば、2-gram (bigram)とは2文字の

つらなりで、3-gram (trigram)であれば、3文字のつらなりである。この定義でのn-gramは、Biber and Barbieri (2007)の言った multi-word sequences (複数語のつらなり)と通じている。ほかに lexical phrases、formulas、routines、fixed expressions、pre-fabricated patternsといった呼び方もある。Biber and Barbieriによると、複数語のつらなりの意味本質はいくつかの側面から観察することができる(2007: 264)。語句が表面的に示す意味ではなく、特定の意味に用いられる慣用句(multi-word sequences that are idiomatic)に注目する研究もあれば、慣用句ではないが、語と語の結びつきが顕著である語句(sequences that are non-idiomatic but perceptually salient)に注目する研究もある。また、Biber and Conrad (1999)、Biber and Barbieri (2007)は、高頻度に現れる語のつらなり(the most frequently recurring sequences of words)、たとえば *I don't know if* や *I just wanted to* などのようものに関心を寄せている。このようなつらなりをlexical bundles (語の束)と呼んでいる。

n-gram は文体やスタイル研究、あるいは文章の書き手の特定に用いられている。石井(2001)はn-gram の手法によって、仏教文献の異本を確認し、文章の書き手や翻訳者の特定に有効であると検証している。近藤(2000)は、『古今和歌集』より抽出した n-gram とコロケーションによって、和歌の書き手が男性か女性かを推定することに成功した。これによって、書き手の性別は和歌の言語特徴に大きくかかわっていることが明らかになった。ほかに、松村・金田(2000)、金明哲(2002)などでも、書き手の特徴を抽出する際にn-gram が有効だと明らかにされている。このように、高頻度に使われる語や文字のつらなりやかたまりは書き手の特徴・癖として捉えることができる。

田野村(2011)は n-gram を用いて、芥川龍之介と太宰治の作品の中から、非常に類似した表現を発見している。a、bはその例である<sup>2)</sup>。

- a. が、ともかくもこの麻利耶観音には、気味の悪い因縁があるのだそうです。(芥川龍之介 1920『黒衣聖母』)
- b. どうもあまり、大人げのない話でございますが、その橋には気味の悪い因縁があるのださうで、もともとその橋は・・・(太宰治 1943『右大臣実朝』)

aとbのような、文の主題に違いがあるほか、互いに似通った表現の例がいくつか見つかった。作品の年代から見れば、太宰治が芥川龍之介に影響されている可能性があることが示唆されているが、二人の間にはどのようなつながりがあるかはこれらの事例だけでは説明できない。n-gram の手法は確かに言葉がどのように使われている、あるいはどのような特徴があるかといった証拠を提示することができる。しかし、それを出発点として、どのようにその背後にある因果関係を証明するためには、さらに作家個人の生い立ちや執筆背景などを掘り下げていく必要があると思われる。

翻訳研究においては、翻訳者の文体(スタイル)が理論的に論じられてきている。Hermans (1996)によると、訳者の「ヴォイス」が翻訳テキストにはあらゆる文字の中に表れる(“the other voice is there in the text itself, in every word of it” p.9)。翻訳テキストへの付加、あるいは訳文の注釈など訳者が一人称の立場で行う説明などは、翻訳者の「ヴォイス」に当たる。「ヴォイス」は読者がその存在に気づかないようにテキストの中に隠されているだけであると Hermans は言う。

Baker(2000)は二人の訳者による翻訳作品から3-gram、4-gram を抽出し、その使用頻度から翻訳者の

言語使用傾向を分析し、Hermans (1996) の言う、読者が見えない翻訳者のスタイルを明らかにする実証的な研究を行った。Baker (2000:245) によると、翻訳者のスタイルとは、言語的あるいは非言語的な指紋であり、たとえば、どんなタイプの原文をよく翻訳するのか、あるいは訳注や前書きなどでも見られる翻訳者が持続的に使用する方略などである。これらは、翻訳者個人の典型的な表現であり、一時的な原文の影響による使用ではないということを確認するのも重要である。Baker は特に翻訳者の言語表現—語彙、統語法、または接続語や句読点などの記号がパターンとなるものに関心を持っている。したがって、パターンとなる言語的な特徴を明らかにするには、コンピューターによる分析方法が有効だと思われる。

人間であればだれでも個人の言語使用習慣があり、訳者も例外ではない。訳者は自らの言語表現の個性や原作者の言語表現の習慣の間に良いバランスをとるようにつとめるが、意図的、あるいは非意図的に、個人の特徴は多かれ少なかれ残される。本研究ノートは、これを基礎仮説として、同一人物が翻訳したテキストにおける語彙使用の特徴および分布状況を分析・観察していきたい。

### 3. コーパスの構成

本研究の分析データは日本語の原文の小説4冊、および劉慕沙による中国語の翻訳本から構成される。安本(1965)の文章の特徴についての分類に基づき、劉慕沙の50作あまりの翻訳作品とその日本語の原文の小説の中から、異なる文体類型を示す三島由紀夫、川端康成、石川達三、大江健三郎の4人の作品を選んだ<sup>3)</sup>。

日本語小説ならびに劉慕沙の中国語翻訳本の詳細は以下の通りである。中国語・日本語の作品をより分かりやすく示すため、以下では日本語小説の場合は『』で、中国語翻訳本の場合は《》でくくることにする。

#### 日本語原文

三島由紀夫(1953)『潮騒』東京:新潮社

川端康成(1955)『女であること』東京:新潮社

石川達三(1948)『幸せの限界』東京:新潮社

大江健三郎(2000)『取替え子』東京:講談社

#### 劉慕沙による以上の作品の中国語翻訳

《潮騒》(1970)台北:阿波羅出版社(初版)<sup>4)</sup>

《女身》(1984)台北:三三書坊(初版)

《幸福的限界》(1991)台北:遠流出版社(初版)

《換取的孩子》(2002)台北:時報文化出版社(初版)

以上の8点の作品をパソコンで読み込めるようにすべて文字化し、Wordsmith (Ver. 6.0) のwordlist clusterの機能を使用して、n-gramを抽出した。なお、以上のBaker (2004) の訳者の特徴や松村・金田(2000)、金明哲(2002)など著者推定の研究とは異なり、本論文では一人の翻訳者の作品を研究対象としているため、特定訳者の言語使用の特徴であるかどうかを判定する際には、参照コーパスの台湾中央研

究院が公開した500万語の中国語Sinica Corpus (1994年公開) および北京日本学研究中心によって開発された「中日対訳コーパスCJCS」第一版(2003年発表)を用いている。

#### 4. データと事例分析

##### 4.1 データ:n-gram の分布

##### 4.1.1 中国語翻訳作品における n-gram の分布

本節では、翻訳文における隣接した文字の n-gram から、翻訳者個人の言語使用のパターンを探る。まず、中国語翻訳作品において頻度 5 以上の 3-gram から 10-gram のタイプ数を図 1 に、4 つの翻訳作品の文字数を表 1 に示す。

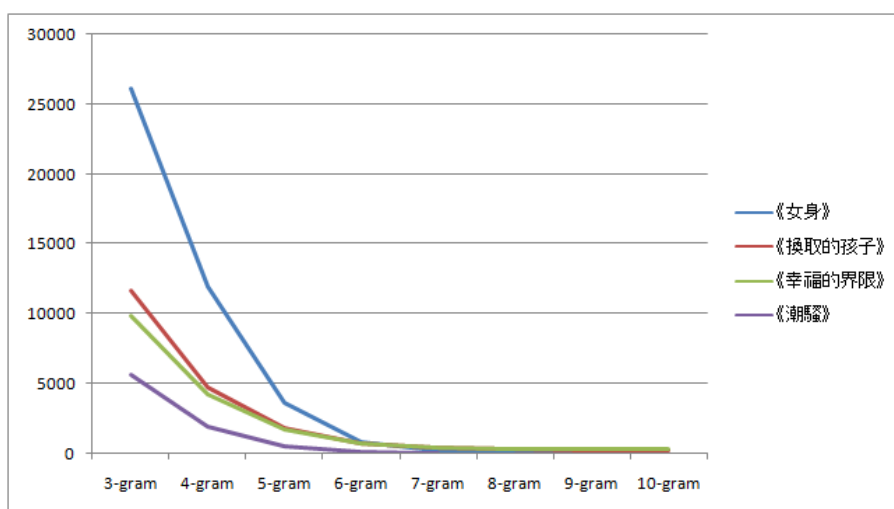


図 1. 4 つの翻訳作品の 3-gram から 10-gram までのタイプ数

表 1. 4 つの翻訳作品の文字数

	《女身》	《換取的孩子》	《幸福的界限》	《潮騷》
文字数	211,992	134,452	98,414	68,545

表 1 と図 1 を照らし合わせてみると、n-gram はもともとのテキストの大きさに左右されることがわかる。《女身》は文字数が最も多いため、その n-gram のタイプも多くあり、逆に《潮騷》は 4 つの作品の中に文字数が最も少ないため、n-gram のタイプも比較的到低い傾向にある。

さらに、それぞれの作品の中に 3-gram から 10-gram までもっとも頻度の高い上位 20 位の文字列の内訳を以下の表 2 に示す。それぞれの作品に登場する主人公の名前や内容に関する言葉が上位に上がっている。また、時間節をあらわす「的時侯」(〜とき)はどの作品にも高頻度に使用されている。このように、各作品の高頻度 n-gram 文字列は原文の内容に関わっていると思われる。

表 2. 作品ごとの高頻度の n-gram 文字列

	《女身》	《換取的孩子》	《幸福的界限》	《潮騷》
上位 20 位	1. 的時候(203)	古義人(1051)	由岐子(295)	千代子(57)
	2. 千代子(147)	大黃哥(124)	子夫人(182)	管理員(55)
	3. 的事情(145)	吾良的(96)	敦子夫(144)	的時候(50)
	4. 榮子的(141)	的吾良(87)	敦子夫人(144)	新治的(49)
	5. 是不是(132)	自己的(69)	自己的(112)	兩個人(43)
	6. 市子的(121)	義人的(56)	的生活(74)	初江的(38)
	7. 這樣的(113)	古義人的(56)	怎麼樣(59)	的影子(36)
	8. 的榮子(111)	是吾良(55)	女兒的(52)	的事情(36)
	9. 兩個人(102)	的時候(52)	有什麼(51)	的新治(34)
	10. 一個人(94)	的古義(48)	丈夫的(48)	村子裡(33)
	11. 妙子的(86)	的古義人(48)	的時候(47)	的母親(33)
	12. 了市子(83)	的東西(48)	的幸福(46)	自己的(31)
	13. 有什麼(82)	小伙子(43)	一個人(44)	的地方(27)
	14. 自己的(80)	的電影(43)	這樣的(43)	暴風雨(27)
	15. 不知道(79)	是古義(42)	母親的(41)	少女的(26)
	16. 佐山的(77)	是古義人(42)	麼樣的(41)	的青年(26)
	17. 怎麼樣(76)	浦小姐(40)	夫人的(40)	也沒有(26)
	18. 麼樣的(74)	吾良兄(39)	生活的(37)	三個人(24)
	19. 也沒有(72)	吾良先(37)	沒什麼(37)	這樣的(24)
	20. 了榮子(72)	吾良先生(35)	是不是(35)	來新治(22)

注: 括弧内の数字は出現頻度。

#### 4.1.2 中国語翻訳作品に共通した n-gram の文字列

翻訳作品にわたって、高頻度に使用されている言語項目を見出すために、4 つの中国語翻訳小説において、頻度 5 以上、且つ 3 作品以上に現れる n-gram の文字列を抽出することにした。また、n が小さい文字列は、意味不明のゴミが入っていることを考量し、ここでは 4-gram、5-gram、6-gram、7-gram の文字列のみを調べた<sup>5)</sup>。表 3 は、抽出結果の内訳である。

表 3. 中国語翻訳小説に現れる 4-gram、5-gram、6-gram、7-gram (頻度 5 且つ 3 作品以上)

	4-gram	5-gram	6-gram	7-gram
タイプ	872	122	10	1
上位 25 位	1. 也不曉得(59)	意想不到的(20)	不曉得是不是(10)	到底是怎麼回事 (5)
	2. 什麼樣的(59)	不管怎麼樣(18)	發生了什麼事(7)	
	3. 沒有什麼(53)	是怎麼回事(18)	定定的凝視著(6)	
	4. 做母親的(49)	若無其事的(18)	麼樣的一個人(6)	
	5. 那麼樣的(40)	也沒有什麼(17)	一切的一切都(5)	
	6. 回過頭來(39)	一動不動的(16)	到底是怎麼回(5)	
	7. 怎麼回事(39)	一點兒也不(15)	底是怎麼回事(5)	
	8. 那個時候(38)	是什麼樣的(15)	是不折不扣的(5)	
	9. 這麼一來(38)	倒不如說是(15)	種無以言喻的(5)	
	10. 這麼樣的(34)	也不曉得是(14)	頭也不回的走(5)	
	11. 無論如何(34)	不知為什麼(13)		
	12. 來的時候(31)	在榻榻米上(13)		
	13. 一開始就(30)	的那個時候(13)		
	14. 什麼時候(29)	第二天早晨(13)		
	15. 會不會是(29)	一心一意的(12)		
	16. 不一會兒(28)	一本正經的(12)		
	17. 在一起的(27)	一廂情願的(12)		
	18. 做丈夫的(27)	是這麼樣的(12)		
	19. 就是這樣(27)	從一開始就(12)		
	20. 今天晚上(26)	一切的一切(11)		

注: 下線は 4 作品とも現れたもの。括弧内の数字は出現頻度。

表3から分かるように、n が大きいほど、文字列の数が少なくなる。5-gram は、122 タイプの文字列がある。6-gram の文字列になると、数は一段と下がり、異なり 10 タイプの文字列となった。7-gram になると、「到底是怎麼回事」(いったいどのようなことになっているか/いったいどういうことなのか)の一例しかない。また、4 作品に現れた文字列(下線部分)は 3 作品よりも、数が少なくなり、6-gram と 7-gram とは 0 となった。

これらの文字列は、慣用句からなるもの、たとえば「一本正經」(真面目な様子)、「一廂情願」(相手をかまわず一方的な願望)、結びつきの強い文字列、たとえば「一切的一切」(全部の全部)、「到底是怎麼回事」(いったいどういうことなのか)、「不管怎麼樣」(いくら...ても)などがある。もちろんこれらはすべて Biber and Barbieri (2007) の言った Lexical bundles (語の束)、高頻度に現れる語のつらなりである。

#### 4.1.3 日本語原文作品に共通した n-gram の文字列

参考として、日本語原文作品の n-gram のリストを示しておく。表 4 から分かるように、言語類型が異なる中国語と日本語は n-gram の違いも明らかである。日本語の n-gram のほとんどは、「～なければならぬ」、「～じゃありませんか」、「～ことはなかった」のような定型化した表現である。中国語翻訳の n-gram は、日本語原文の内容に強く影響されているが、日本語原文の n-gram とは関係が薄いようである。

表 4. 日本語小説に現れる 7-gram、8-gram、9-gram (頻度 5 且つ 3 作品以上)

	7-gram	8-gram	9-gram
タイプ	60	11	2
上位 20 位	1. <u>なければならぬ</u> (39) 2. <u>じゃありませんか</u> (17) 3. <u>ことはなかった</u> (12) 4. <u>たのではなかった</u> (10) 5. <u>ることができない</u> (10) 6. <u>たことはなかった</u> (9) 7. <u>ているということ</u> (9) 8. <u>いるのではないか</u> (8) 9. <u>ていたのではない</u> (8) 10. <u>なかったのである</u> (8) 11. <u>になってしまった</u> (8) 12. <u>らなかったしかし</u> (8) 13. <u>うことはなかった</u> (7) 14. <u>かなければならぬ</u> (7) 15. <u>ことができなかつ</u> (7) 16. <u>たのではないかと</u> (7) 17. <u>っているのである</u> (7) 18. <u>ているのではない</u> (7) 19. <u>とができなかつた</u> (7) 20. <u>ねばならなかつた</u> (7)	かなければならぬ(7) ことができなかつた(7) ているのではないか(7) ったことはなかった(6) なければならぬの(6) いうことはなかった(5) うなことはなかった(5) ければならなかつた(5) ったような気がする(5) なければならなかつ(5) ようなことはなかつ(5)	なければならなかつた(5) ようなことはなかつた(5)

注: 下線は 4 作品とも現れたもの。括弧内の数字は出現頻度。

#### 4.2 研究事例: 中国語翻訳作品に見られた特殊な文字列

中国語翻訳作品に共通した n-gram の文字列(表3)から、特殊な文字列を発見した。6-gram の「定定的凝視著(じっと見ている)」は、副詞的な表現の「定定的」(じっと)と、動詞の「凝視」(凝視する、じっと見つめている)に状態の持続を表すアスペクトマーカ―の「著」からなるかたまりである。しかし、一般に使用さ

れている表現とは言い難い。コンコーダンスで「定定」が文章の中でどのように使用されるかをみる。図2は4 翻訳作品から抽出したものである<sup>9)</sup>。

N	Concordance	File
1	沉靜些。少女的眼睛定定的凝視著西邊海上	sio.word.txt
2	去看這個青年，而仍舊定定的停留在遠海那邊。	sio.word.txt
3	母親的洗著餐具，一面定定的望著再度走進	sio.word.txt
4	他笑笑。安夫卻板著臉定定的望著這邊，接著	sio.word.txt
5	淚濕的面頰，幽暗裡定定的俯視著她白皙的	siawase.word.txt
6	了，轉過頭去，卻見她定定的凝視著天花板。「	siawase.word.txt
7	」省子沒有回答。夫人定定的凝視著女兒的背影	siawase.word.txt
8	車窗外的風景說。夫人定定的凝視著那父女倆，	siawase.word.txt
9	原就無所謂新舊。夫人定定的凝望著預備踏上	siawase.word.txt
10	街上的景色，而自管定定的凝視著自己的內心	onna.word.txt
11	有這麼久了麼？」清野定定的凝視著市子：「	onna.word.txt
12	的動靜使她回過頭來，定定的望著他，待他	onna.word.txt
13	事情？」「是的。」妙子定定的望著河邊，眼看著	onna.word.txt
14	的。」市子睜開眼睛，定定的凝視佐山的臉。	onna.word.txt
15	，可是總會留下什麼，她定定的站在那裡，一再	onna.word.txt
16	步的說。「唔？」佐山定定的瞪著白色的天花板	onna.word.txt
17	在那兒作畫。果然，她定定凝視著素描簿，說出	changing.word.txt

図2. 翻訳作品における「定定」のコンコーダンス

図2によると、「定定」は全部で17例あり、4作品とも使用されており、また視覚動詞と一緒に使用することが多いことが分かる。パラレルコーパスから「定定」が含まれる文脈とその日本語原文を抽出し、原文による影響をしてみる。下の例(1)のように体が固まって動いていない状態の使い方以外に、その他の16の例はある物体に集中して見ている状態という意味で使われている。その(2)~(4)のように、視覚動詞の「凝視」(見つめる)、「俯視」(鳥瞰する)、「瞪」(目を大きくして見つめる)、「望」(眺める)と一緒に使用されている。その日本語の原文と照らしてみると、対応する日本語原文には副詞の「じっと」、「しっかり」があることが分かった。

- (1) 妙子很有條理，總喜歡把東西收拾得整齊清潔，她本身用的東西也不多，可是總會留下什麼，她定定的站在那裡，一再的環顧四周。《女身》  
妙子はものをきちんと片づける方だし、自分のものも乏しいが、それでもなにかある。あたりを見まわしながら、じっと立っていた。
- (2) 夫人定定的凝望著預備踏上旅途而去的女兒那副背影所流露出來的凜然之氣。《幸福》  
夫人は旅に出て行く娘の後姿の凜々しさをじっと見ている。
- (3) 一雙眼睛並不去看這個青年，而仍舊定定的停留在遠海那邊。《潮騒》  
少女はかろく眉をひきしめた。目は若者のほうを見ずに、じっと沖を見つめたままであった。
- (4) 果然，她定定凝視著素描簿，說出了想必思考多日的一番話。《換取》  
千櫻はやはりスケッチブックをしっかり見つめたまま、それまでずっと考え続けてきたは

ずのことを話した。

実際に中国語の「凝視」には、目を凝らして、よく見ている意味が含まれるが、「定定的」と「凝視著」が一緒に使用すると、意味の重なりになってしまう。一方、視線の集中するさまを強く強調している意味にもなる。訳者の習慣的な表現、特徴的な表現であるかどうかを確かめるため、本研究で扱ったテキスト、中国語辞書、中国語コーパス、日中対訳コーパスなどを用いて探ることにした。

まず、Sinica Corpus を確認したところ、このような「定定的凝視著」の使用ばかりでなく、「定定」を副詞的な機能で使用され例は一つも見当たらない。台湾の教育部辞書、中国大陸のインターネット辞書にもこのような使い方がないため、日常生活で定着して広く使用されている表現ではないと言える。

さらに、参照コーパスとして、北京中日対訳コーパスで「定定」を調べたところ、中国語原文の場合では8例、4人の作家に集中していることがわかった<sup>7)</sup>。しかし、中国語の翻訳文では1例しか見当たらない。これらの例文では、「定定的」は「看」(見る)、「望」(眺める)、「瞧」(見る)などの視覚動詞と共起しているが、「凝視」と共起する例はない。「定定」は小説などの文学的な表現として使用されていると言える。だが、「定定」と「凝視」が共起する用法はかなり特徴的だと言えよう。

表4にはのせられなかったが、4-gram のリストから、「久久、久久」という4文字のかたまりを発見した。ここの「久久、久久」は、「久久」を2回重複した表現であり、4つの翻訳作品から6回使用されている。普通、「久久」は例(5)のように副詞として述語の前に置かれ、「長い間」という意味で用いられる。(6)~(9)における「久久、久久」は、文頭や文末に置かれ、文のその他の部分とコンマで隔てている。ここで、「久久、久久」は述語を修飾する副詞の機能よりも、文全体の副詞である役割を担うことになる。つまり、視覚、聴覚、思考などの動作が長い間持続するということを強調するだけでなく、文全体で描かれた映像が一つの静止画面となり、動作や仕草が長らく持続していることを示すことで、読者により強い印象を与える効果がある。

- (5) 然而，市子還是久久無法成眠。她止不住去想像榮子伸展四肢，悠然安睡在那張單人床上的模樣。《女身》

市子は長いこと寝つけなかった。さかえが一人のベッドに、のびのびと手足をのばしているのを思ってみたりした。

- (6) 浦小姐離去後，趁著古義人與小明從俱樂部回來前，攤開仙達克的《外邊的那一頭》，凝視著愛姐為尋找小妹妹一開始以錯誤姿勢飛出窗外的畫面，久久，久久。千樑自己也必須慎重採取正確的行動才好。《換取》

ひとりになった千樑は、古義人とアカリがクラブから帰って来る前に、センダックの“Outside Over There”を拵げると、妹を探しにアイダが夜の窓の外へ飛び出して行く、その最初の、過った姿勢をとってしまうシーンを、長い間みつめていた。千樑もまた、慎重に正しく行動しなければならないのだった。

- (7) 而榮子，平靜是恢復了，卻把光一遞給她的燒賣放在膝蓋上，將那張微垂的面孔對著窗外，久久，久久。《女身》

さかえも落ちつきを、取りもどしてはいた。しかし、光一から渡された、焼売をひざにお



いたまま、うつ向き気味の顔を、窓のそとに向けていた。

- (8) 母親凝視著耀眼的天空，直到那隻蝴蝶化為一個黑點。久久，久久，蝴蝶始終在她視野的一角振翼飛翔，末了，也許是絕望於大海的浩瀚與閃爍的眩惑之下，鄰島那看似很近、飛起來卻又那麼長遠的距離，牠終於貼近海面，搖搖盪盪的飛回到堤堰上來。《潮騷》

母親は蝶が黒い一点になるまで眩い空をみつめた。いつまでも蝶は視界の一角に羽ばたいていたが、海のひろさと燦めきに眩惑され、おそらくその目に映っていた隣りの島影の、近そうで遠い距離に絶望して、今度は低く海の上をたゆたいながら突堤まで戻って来た。

- (9) 在擁擠的電車裡，她抓著吊環，閉上眼睛，久久，久久。《幸福》

彼女は混雑した電車のつり革につかまって永いあいだ目を閉じていた。

ここでも、Sinica Corpus、中国語のインターネット辞書を調べた。その結果、「久久、久久」は一例もないという結果となった。ただし、北京中日対訳コーパスでは、中国語翻訳テキストと原文テキストがそれぞれ 1 例あり、文学作品の中での使用であることが分かる。その内、例(10)は主語と述語の間に置かれ、副詞の用法である。例(11)は文の真ん中にあるが、後の文節を修飾する機能は劉の用法に近い。使用頻度と文中の位置から考えると、文末や文頭に置かれる「久久、久久」の使用は、劉の独特な言語使用の特徴と言えよう。

- (10) 我久久、久久地等待着。(CJCS 『挪威的森林』 翻訳文)

僕はいつまでも待ちつづけた。(CJCS 『ノルウェイの森』 原文)

- (11) 信，念完了，久久，久久，屋里没有一点声音。(CJCS 『轮椅上的梦』 原文)

手紙は終わった。しばらく部屋の中は何の物音もしなかった。(CJCS 『車椅子の上の夢』 翻訳文)

ちなみに、劉が 2002 年に出版した《甘露》(吉本ばなな(1995)『アムリタ』角川文庫)、および 1970 年に出版した《美德的動搖》(三島由紀夫(1957)『美德のよろめき』大日本雄弁会講談社)においても、「定定的凝視著」と「久久、久久」の表現が確認された。劉の作品をすべて調査したわけではないが、ここで言えることは、翻訳者はあるコンテキストにおいて習慣的に使用する表現があり、これが個人の言語使用の痕跡として翻訳テキストに残ったのであろう。

## 5. 考察と今後の課題

本研究は n-gram を用いて、翻訳者劉慕沙が翻訳テキストに残した言語使用の特徴(痕跡)の発見を試みた。「定定的凝視著」と「久久、久久」の表現は、調査した 4 作品にわたって使用されている上に、その他の参考コーパスでもめったに見られない。したがって、訳者には独特な語彙使用の習慣があると確認することができた。興味深いことに、《潮騷》(1970)と《取り換え子》(2002)は 30 年以上の歳月も隔てているにもかかわらず、二つの表現が長い間使用されているのである。これは、単なる独立した使用パターンなのか、それとも翻訳者個人のスタイルの現れなのかは、まだ不明である。後者の翻訳者個人のスタイルの現れというのは、たとえば、劉慕沙による重畳形の副詞の使用に特徴的である、あるいは文末文頭の

副詞の使用に独特なパターンがあるということが考えられる。今後の課題として、より多くの翻訳者による翻訳データを集め、個人のスタイルを形作る特徴を検証していきたい。

また、本研究の研究方法は改善すべき点が多い。パソコンによってあらゆる隣接した文字の連続を引っ張り出した後、数多くの文字列から研究者の直感を頼りに特殊な表現を見出す、といった手順を踏んでいかねばならないため、システム的で、網羅的な方法とは言えない。人手に頼らずより効率的な方法を築く必要があると思われる。さらに、パターンとなる構文や統語などの特徴の検出方法の開発も必要だとと思われる。

以上、本研究ノートは分析結果の解釈も未解決の状態にとどまり、研究方法も未熟であるが、ひとまず分析データと一部の結果をまとめて、これからの研究展開のきっかけとしたい。

.....  
**【著者紹介】**

鄧敏君(Teng, Minchun)、台湾致理技術学院応用日本語学科助理教授。コーパス言語学の視点から、翻訳・通訳を研究している。連絡先:E-mail: claireteng@hotmail.com

.....

**【註】**

1. 隣接している文字や語だけが **n-gram** になるとは限らない。金明哲(2002)は、日本語の助詞の **n-gram** モデルに基づき、日記文の書き手の識別を行った。助詞の **n-gram** モデルとは、文中の助詞以外の単語を無視し、隣接している助詞のみを分析対象とする分析方法である。
2. 傍点と太字は筆者。
3. 安本(1965)は因子分析の方法を用いて、100 篇の作品を調査した。安本(1965)によれば、三島由紀夫は「体言型—非修飾型—文章型」に、川端康成は「用言型—修飾型—会話型」に、石川達三は「体言型—非修飾型—会話型」に、大江健三郎「体言型—修飾型—文章型」の文体類型に分類されている。なお、本文は安本・本田(1981)と安本(1994)も参考している。
4. 《潮騒》は、1970 年に阿波羅によって出版されたが、1991 年には遠流出版社によって再出版され、内容はほぼ変更されていない。《女身》もそれぞれ 1984 年に三三書坊、1991 年に遠流出版社によって出版されたものがある。
5. 頻度 5 以上、且つ 3 作品以上の中国語翻訳小説に現れるという条件に合致する **8-gram** 以上の文字列は存在しない。
6. 表 3 では、「定定的凝視著」が 3 作品の中に使われている。しかし、もう一つの作品では「定定凝視著」を使っている。つまり、4 作品とも使用されていると言える。
7. 抽出した例文を 3 例あげておく。
  - a. 他和随随送我到县城。娃娃们追过河，追着我们的驴车跑，终于追不上了。就都站下来定定地望着我们走远。『插队的故事(原文)』
  - b. 她每次称完草，总要用那双大眼睛定定地看着我，好象要从我的眼睛里看出些什么。『轮椅上的梦(原文)』

- c. 猛地，她象松了劲儿似的瘫软了，圆睁的眼睛定定地呆住了，两只悬在胸前的手一下摔在了炕席上。  
『轮椅上的梦（原文）』

### 【参考文献】

- Baker, M. (1993). Corpus Linguistics and Translation Studies: Implications and Applications. In M. Baker (Ed.) *Text and Technology: In Honor of John Sinclair*. (pp.231-250). Philadelphia/Amsterdam: John Benjamins Publishing Company.
- Baker, M. (2000). Towards a Methodology for Investigating the Linguistic Behaviour of Professional Translators. *Target*. 12 (2). pp. 241-266.
- Biber, D. and Conrad, S. (1999). Lexical Bundles in Conversation and Academic Prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of Corpora: Studies in Honor of Stig Johansson* (pp. 181–189). Amsterdam: Rodopi.
- Biber, D. and Barbieri, F. (2007). Lexical Bundles in University Spoken and Written Registers. *English for Specific Purposes*. 26. pp.263-286.
- Saldanha, G. (2011a). Translator Style: Methodological Considerations. *The Translator: Studies in Intercultural Communication*. 17(1). pp.25-50.
- Saldanha, G. (2011b). Style of Translation: The Use of Foreign Words in Translation by Margaret Jull Costa and Peter Bush. *The Corpus-based Translation Studies: Research and Applications*. pp.237-258.
- Hermans, T. (1996). The Translator's Voice in Translated Narrative. *Target*. 8:1. pp.23-48.
- 石井公成(2001)「N-gram 利用の可能性—仏教文献における異本比較と訳者・作者判定—」『漢字文献情報処理研究』(2) pp.59-61
- 金明哲(2002)「助詞の n-gram モデルに基づいた書き手の識別」『計量国語学』23 巻 5 号 pp.225-240
- 近藤みゆき(2000)「nグラム統計処理を用いた文字列分析による日本古典文学の研究—古今和歌集のことばの型と性差—」『人文研究』第 29 号 pp.187-238
- 田野村忠温(2011)「コーパス言語学の新たな展開」『日本語学』第 30 巻 14 号(2011 年 11 月臨時増刊号『言語研究の新たな展開』) 明治書院 pp.95-104
- 松浦司・金田康正(2000)「n-gram の分布を利用した近代日本語の著者推定」『計量国語学』22 巻 6 号 pp.225-238
- 安本美典(1965)『文章心理学入門』誠信書房
- 安本美典(1994)「文体を決める三つの因子」『月刊言語』23 巻 2 号 pp.22-29
- 安本美典・本多正久(1981)『因子分析法』培風館

