

텍스트 마이닝 기법을 활용한 번역가 문체 분석*

- ‘as if’ 구문 번역을 중심으로 -

신혜정** · 문성민***

Hyejung Shin and Seongmin Mun (2022), A Text Mining Analysis of Translator's Style: Comparing Translating Styles of “As If” Constructions. *This study compares the translating styles of “as if” constructions in multiple translator corpora, using text mining techniques. First, hierarchical cluster analysis and t-stochastic neighbor embedding (t-SNE) techniques were applied to the most common words in translations of “as if” constructions from a corpus of 11 English-Korean translations by three translators. The cluster analysis and t-SNE results established the presence of relative similarity/dissimilarity between each translator's works by visually presenting the relative distance between the texts. To further analyze which text is similar or dissimilar to which and to verify similarities in translations by the same individual translator, the study followed up with a chi-square test and correlation analysis. This paper then discusses translator-specific patterns in detail, accompanied by examples from each translator's texts. The results show that a multi-layered, quantitative analysis of translation patterns of “as if” constructions is a valid and effective method to identify a translator's style in literary translation. (Chosun University, Korea)*

Keywords: translator's style, “as if” constructions, function words, stylometry, t-SNE

주제어: 번역가 문체, ‘as if’ 구문, 기능어, 계량적 문체 분석, t-분포 확률적 임베딩

* 본 논문은 조선대학교 연구지원금의 지원을 받아 연구되었음(2020학년도).

본 논문은 신혜정(2021)의 내용에 분석 방법 및 분석 대상을 추가·보완하여 발전시킨 연구임.

** 조선대학교 영어영문학과, 부교수, 제1저자

*** 조선대학교 영어영문학과, 박사후연구원, 공동저자

1. 서론

텍스트 마이닝(text mining)은 문자로 된 데이터를 균일한 형식으로 정제하여 텍스트 데이터에 내포된 특징을 추출하고 새로운 정보를 찾아내는 기법이다. 여기서 마이닝(mining)이란 광산에서 광물을 캐내듯이 다량의 데이터에 숨겨진 패턴과 연관관계를 찾아 유의미한 정보로 추출한다는 의미이다. 텍스트 마이닝은 크게 자료 처리 과정과 자료 분석 과정으로 나뉜다. 자료 처리 과정에는 사용자가 필요한 정보와 관련된 텍스트가 들어 있는 특정 문서를 찾고 그 특정 문서에서 구체적인 정보를 정제하는 과정, 그리고 인간의 언어인 자연어를 컴퓨터로 분석하기 위해 가공하는 과정이 포함된다(Hotho et al. 2005: 4). 이러한 자료 처리 과정을 거친 후 자료 분석 과정에서 데이터 마이닝, 기계학습(machine learning), 통계학 등을 활용하여 의미 있는 결과를 도출한다.

정해진 구조가 없는 텍스트, 즉 비정형 데이터(unstructured data)인 문학작품도 컴퓨터로 자료를 처리하는 정량분석 방법을 활용하여, 인간의 자세히 읽기 및 해석만으로 알아내기 힘든 유의미한 정보를 찾을 수 있다. 소설과 같은 문학작품은 정해진 형식이 없는 텍스트이므로 이를 양적으로 분석할 때 주로 탐색적(exploratory) 분석 방식을 적용하여 텍스트의 특징을 찾는다.¹⁾ 먼저 텍스트들을 모아 전자화한 후, 정량화를 위해 데이터 처리 작업을 거친다. 이후 데이터를 탐색하여 텍스트 패턴을 찾는다. 이러한 텍스트 패턴들을 종합, 분석하여 새로운 정보를 얻는다. 문체는 “언어사용에서 나타나는 특징”(Leech and Short 2007: 10)을 말하므로 텍스트 마이닝의 탐색적 분석 기법은 저자의 문체 포착에 용이하다. 또한 동일한 기법을 번역가의 문체 분석에 적용할 수 있다.

코퍼스를 활용하여 번역가 문체를 연구한 베이커(Baker 2000)는 ‘번역가의 반복되는 언어사용 패턴’을 번역가 문체라고 정의하면서 특정 번역가의 번역 텍스트들에서 공통으로 나타나는 패턴을 연구하였다. 그러면서 그러한 패턴이 다른 번역가와 해당 번역가를 구별하는 해당 번역가만의 문체인지, 즉 ‘번역가 지문(thumb-print)’(Baker 2000: 245)인지 파악하기 위해 동일 원문(source text: ST)의 복수 번역문(target text: TT)도 함께 분석할 필요가 있다고 하였다. 베이커(Baker 2000)와 마찬가지로 TT 중심의 번역가 문체를 연구한 살다냐(Saldanha 2011)도

1) 코퍼스 주도(corpus-driven) 접근법으로 한영 번역소설을 분석한 이창수(2015; 2020)와 최희정(2020)도 이러한 탐색적 분석 방식을 취해 번역 문체를 연구하였다.

특정 번역가가 번역한 TT들을 모은 번역가코퍼스를 분석함과 동시에, 해당 번역가를 다른 번역가와 구별해주는 일관된 패턴을 발견하기 위해 동일 ST의 복수 TT를 분석해야 한다고 강조했다. 이렇듯 TT 간 차이만이 아니라 개별 번역가의 문체를 논하기 위해서는 동일 ST의 복수 TT 코퍼스와, 동일 번역가들의 여러 TT를 모은 개별 번역가코퍼스 모두를 분석해야 한다. 그러나 지금까지 국내 번역가 문체 연구에서 양쪽 코퍼스 모두를 동시에 분석한 연구는 찾기 어려웠다. 대부분의 선행연구가 동일 ST에 대한 2개의 TT를 비교한 것이거나, 번역가도 각기 다르고 원저자도 겹치지 않는 여러 번역 텍스트를 모아 분석한 것에 그쳐 개별 번역가 문체를 논할 수 있는 조건을 충족하지 못하였다.

이에 본 연구는 동일 ST의 복수 TT 분석(1차)과 동시에 개별 번역가코퍼스 분석(2차)을 진행하여 번역가 문체를 밝히고자 한다. 영어 ‘as if’ 구문을 중심으로 세 번역가의 영한 번역 데이터를 텍스트 마이닝 기법으로 분석하여 각 번역가의 문체 특징을 밝히고, ST와 TT의 텍스트 예문을 분석하여 번역가 간 문체 유사성과 차이점을 논의한다.

2. 분석 대상 및 분석 방법

2.1. ‘as if’ 문체 표지

문체를 가리켜 “특정 텍스트에 나타나는 언어 특징”(Leech and Short 2007: 11)이라고 한 리치와 쇼트는 소설가 찰스 디킨스(Charles Dickens)의 문체 표지(style marker)로 ‘as if’를 예로 들어 ‘문체’를 설명한다(Leech and Short 2007: 51). 리치와 쇼트는 문체를 관찰하기 위해서는 특정 텍스트에서 꾸준하게, 일관된 형태로 나타나는 현상을 관찰해야 한다고 하면서 그 현상이 얼마나 자주, 반복적으로 나오는가 하는 ‘빈도’가 중요하다고 하였다(Leech and Short 2007: 34). 본 연구에서 분석한 텍스트 중에는 디킨스의 작품이 많고 디킨스 외 작가의 작품에서도 ‘as if’가 반복적으로 쓰였다. 또한 코퍼스 문체론(corpus stylistics)으로 디킨스 문체를 연구한 말베르크(Mahlberg 2007; 2013)도 저자 문체 분석 자질로 ‘as if’ 구문 등 엔그램(n-gram)²⁾을 분석하였다. 이창수(2020)도 말베르크 문헌을 인용하여, 문체 연구에서 작가 문체 특징을 드러내는 표현으로 ‘as if’가 자주 분석된다고 하였다.

또한 이창수(2020: 133)는 황순원의 『나무들 비탈에 서다』의 두 영역본의 문체를 분석한 연구에서 번역가 장왕록(Chang Wang-rok)의 문체 특징으로 ‘as if’ 패턴을 확인하고, 두 번역가의 ‘as if’ 사용 차이를 분석하는 것을 후속 연구로 제안하기도 하였다.

‘as if’는 텍스트 내에서 발생빈도가 높은 고빈도어이면서 동시에 내용어를 연결해주는 문법적인 기능을 하는 기능어(function word)이다. 이창수(2020: 119)는 문장 길이, 타입-토큰 비율, 단어 수 등의 다양한 저자 판별(authorship attribution) 자질 중에서 언어에 무관하게 분석 가능하며 가장 성공적인 분석 결과를 보이는 자질로 최빈도어휘(most frequent words)를 꼽았다. 그러면서 텍스트에서 높은 빈도로 쓰이는 단어들이 바로 전치사, 대명사, 조동사 같은 ‘기능어’라고 하였다. 이에 본 연구는 언어에 무관하고 ST 내용과 무관하며 ST 저자와도 무관하게 분석 가능한 문체판별 자질로서 문체 표지이면서 고빈도 기능어인 ‘as if’를 분석 자질로 선택하였다.

서양에서는 저자 판별 연구의 역사가 길어 영어 저작물을 대상으로 한 연구가 활발하지만, 국내의 저자 판별 연구는 시작 단계에 불과하다(김일환과 이도길 2015). 본 연구는 한국어 텍스트, 특히 영한 번역 텍스트에서 ‘as if’의 번역을 대상으로 한국어 기능어의 문체 변별력을 분석하고자 한다. 영어 저자 판별 연구와 마찬가지로 기능어를 분석하여, 교착어인 한국어에서 조사, 어미 등의 문법요소가 번역가 문체 변별자질로서 타당한가를 검증하는 효과도 기대하였다.

2.2. 분석 코퍼스

ST가 같은 복수 TT를 비교하면 ST 변수가 통제되어 (원저자 문체의 영향 없이) 개별 번역가의 문체를 볼 수 있다. 같은 내용을 번역할 때 번역가 사이에 어휘(표현) 사용에서 차이가 난다면 번역가들이 각자 다른 선택을 한 것, 즉 문체가 다른 것을 의미하기 때문이다(Munday 2007: 20). 이렇게 ST 문체를 통제할 수 있다는 점에서 동일 ST의 복수 TT 분석은 번역가 문체 연구에서 필수적이다.

본 연구는 베이커(Baker 2000)와 살다냐(Saldanha 2011)의 번역가 문체 연구에서 착안하여, 분석 대상 텍스트를 선정할 때 다음 조건을 고려하였다. 첫째, 여러

2) n개의 단어가 연속되어 쓰이는 것을 말한다. 본 연구의 분석 대상인 ‘as if’는 2그램이다.

작품을 번역한 번역가를 대상으로 한다. 둘째, 동일 ST를 번역한 복수 TT가 있는 작품을 대상으로 한다. 여기서 중요한 점은 가능한 한 ST 저자 그리고 TT 번역가가 각각 겹치도록 ST 코퍼스와 TT 코퍼스를 구성하는 것이다.³⁾ 이렇게 하면 ST 저자 변수와 ST 주제 변수가 통제되는 효과가 있다. 즉, ST 저자 문체나 ST 내용과 같은 번역가 외부 요인의 영향을 최소화하여 번역가 고유의 문체 식별력을 높일 수 있다. 특히 본 연구는 대상 번역가 모두가 공통으로 번역한 작품을 분석에 포함하여, ST 변수를 통제된 상태에서 개별 번역가 문체를 직접 비교할 수 있도록 하였다. 이러한 코퍼스 설계는 특정 텍스트의 번역 문체가 아닌, 특정 번역가의 ‘번역가 문체’를 밝혀내기 위한 것이다. 살다냐(Saldanha 2011)도 ‘번역 텍스트의 문체(the style of the *text*)’와 ‘번역가의 문체(the style of the *translator*)’를 구별하면서 베이커와 자신은 후자에 관심이 있음을 밝혔다. 본 연구는 이들 요건을 모두 충족하는 세 번역가의 번역코퍼스를 텍스트 마이닝 기법 및 통계 기법을 활용하여 분석한다.

본 연구에서 분석한 번역코퍼스의 구성은 아래 표와 같다.

표 1. 번역코퍼스 구성

번역가	TT	ST	저자
이인규 Lee	1 위대한 유산(2009)	Great Expectations (1861)	Charles Dickens
	2 올리버 트위스트(2018)	Oliver Twist (1837)	Charles Dickens
	3 채털리 부인의 연인 (2003)	Lady Chatterley's Lover (1928)	D. H. Lawrence
류경희 Ryu	4 위대한 유산(2014)	Great Expectations (1861)	Charles Dickens
	5 제인 에어(2010)	Jane Eyre (1847)	Charlotte Brontë
	6 오만과 편견(2017)	Pride and Prejudice (1813)	Jane Austen
	7 맨스필드 파크(2016)	Mansfield Park (1814)	Jane Austen
김옥수 Kim	8 위대한 유산(2016)	Great Expectations (1861)	Charles Dickens
	9 올리버 트위스트(2016)	Oliver Twist (1837)	Charles Dickens
	10 데이비드 코퍼필드 (2018)	David Copperfield (1849)	Charles Dickens
	11 오만과 편견(2018)	Pride and Prejudice (1813)	Jane Austen

3) 번역가 5명의 34개 번역작품을 분석한 이창수는 코퍼스 구성상 번역가 문체와 저자 문체를 구별하기 어려웠던 한계가 있었음을 지적하면서 후속 연구에서 “번역가와 원저자가 겹치도록”(2015: 133) 코퍼스를 구성할 것을 기약했다.

[표 1]에서 강조 표시한 1번, 4번, 8번 ST는 찰스 디킨스의 *Great Expectations* 로, 세 번역가 공통텍스트이다. Lee와 Kim의 공통텍스트는 2번과 9번 ST인 디킨스의 *Oliver Twist*이고, Ryu와 Kim의 공통텍스트는 6번과 11번 ST인 제인 오스틴(Jane Austen)의 *Pride and Prejudice*이다. 분석의 편의를 위해 각 TT를 코드화하여 간략히 표기하였다. 예를 들어 1번 TT 코드는 번역가, 작품명, 저자 정보를 순서대로 넣은 ‘Lee_GE_Dickens’이다. 4번과 8번 TT 코드는 각각 ‘Ryu_GE_Dickens’와 ‘Kim_GE_Dickens’이다. 이렇게 번역가 Lee의 TT 3개(GE, OT, LCL)와 Ryu의 TT 4개(GE, JE, PP, MP) 그리고 Kim의 TT 4개(GE, OT, DC, PP)의 총 11개 TT를 분석하였다.

2.3. 분석 방법

코퍼스 데이터 분석을 위해 먼저 영어 ST들의 ‘as if’ 구문을 검색하였다. 각 작품의 원문 전자파일은 구텐베르크 프로젝트 ‘Project Gutenberg’ (<https://www.gutenberg.org>)에서 내려받아 사용하였다. 이후 각 인쇄본 TT에서 해당 번역문장을 찾아 전사하여 전자파일로 저장하였다. ST 7개에서 분석 대상으로 총 813회의 ‘as if’ 구문을 추출하고, 이에 대한 세 번역가의 한국어 TT 11개에서 해당 번역 문구 총 1만 6115어절(token)을 분석하였다. 전사 과정에서 빠진 부분이나 오자 등 실수가 없는지 확인하고 수정하였다. 이렇게 번역가 Lee, Ryu, Kim의 TT들에 나타난 ‘as if’ 번역문을 입력, 저장한 후 통계 및 데이터 분석 프로그램 R을 이용하여 자료 처리 및 분석 과정을 진행하였다.

본 연구는 번역가 간 문체의 유사도 또는 상이한 정도를 측정하는 방법으로 통계 분석을 사용하였다. 먼저 기계학습을 활용한 탐색적 통계 분석법인 군집분석(cluster analysis)과, 다차원의 데이터를 2차원에 축소하여 보여주는 시각화 기법인 t-분포 확률적 임베딩(t-Stochastic Neighbor Embedding: t-SNE)을 실시하였다. 이후 이 결과를 바탕으로 카이제곱 검정 및 상관관계 분석을 수행하였다.

군집분석은 데이터들의 유사성을 측정하여 유사한 데이터끼리 군집으로 묶는 통계 기법이다. 집단을 분류하는 명확한 기준이 없거나 집단의 속성을 모르는 상태에서 대상을 분류할 때 사용한다. 이렇게 분류된 동일 군집 내 데이터 간의 유사점과 서로 다른 군집 간의 차이점을 분석함으로써 데이터에 대한 새로운 통찰을 얻을 수 있다. 본 연구에서는 계층적 군집(hierarchical clustering) 방법으로 군

집분석을 수행하였으며, 데이터들 사이의 유사도를 구하는 코사인 유사도(cosine similarity)와, 데이터들 사이의 거릿값이 작은 데이터들을 군집으로 형성하는 최단 거리 연결법(single linkage method)을 사용하여 수형도(dendrogram)를 생성하였다. 분석에 앞서 각 작품에서 사용한 번역 패턴의 빈도가 다르기 때문에 이를 평균이 0, 표준편차가 1인 표준정규분포를 사용하여 표준화한 뒤 분석을 진행하였다. 군집분석을 위해 R언어에서 제공하는 proxy 패키지를 사용하였으며, 표준화 점수 변환에는 scale 함수를 사용하였다.

t-SNE는 다차원 데이터를 축소하여 2차원 공간에 시각화하는 차원축소기법이다. 복잡한 데이터를 비슷한 구조끼리 정리하여 시각적으로 보여주는 t-SNE는 데이터의 구조를 이해하는 데 도움을 준다. 다변수 또는 다차원으로 구성된 데이터의 차원축소기법으로는 주성분분석(principal component analysis: PCA), 다차원 척도법(classical multidimensional scaling: MDS) 그리고 t-SNE의 세 가지 기법이 주로 사용된다. 이창수(2015)는 번역가 문체 분별을 위해 최빈도어휘를 변수로 하여 주성분분석과 다차원 척도법 등을 실시했다.⁴⁾ 그러나 주성분분석과 다차원 척도법은 차원을 축소하는 과정에서 전체 데이터 사이의 관계 정보를 모두 표현하지 못하고 유사한 데이터들을 중심으로 차원을 축소하기 때문에 상이한 데이터들의 관계를 잘 표현하지 못하는 한계가 있다(Van der Maaten and Hinton 2008). 반면 본 연구에서 사용한 t-SNE는 모든 데이터의 관계를 표현하는 데 있어 앞서 설명한 두 기법보다 성능이 좋은 것이 장점이다. 따라서 본 연구는 번역가 간 번역 문체의 분별을 확인하는 방법으로 t-SNE를 사용하여 연구를 진행하였다. 분석에는 R언어에서 제공하는 Rtsne 패키지를 사용하였다.

다음 장에서는 이러한 통계 분석을 시행한 결과를 제시하고, 분석 결과에서 새롭게 발견한 내용을 논의한다.

3. 분석 결과

현대 국어에서 직유 표시어 ‘-같이’, ‘-듯(이)’, ‘-처럼’이 주로 사용되듯이(하길 종, 2005: 82) 세 번역가도 ‘as if’ 구문을 한국어로 번역할 때 ‘-처럼’, ‘(-)듯*’, ‘(-)같*’을 주로 사용하였다.⁵⁾ 번역가 Lee, Ryu, Kim이 ‘as if’ 구문 번역 시 사용한

4) 주성분분석과 다차원 척도법에 대한 설명은 이창수(2015) 참조

표현은 아래 [표 2]에서 확인할 수 있다. [표 2]는 코퍼스 분석도구 워드스미스(WordSmith)로 추출한 Lee, Ryu, Kim의 GE(『위대한 유산』) ‘as if’ 구문 번역 2그램 목록이다. 최소 발생빈도는 5회로 설정하였다. 지면 제약상 전체 번역코퍼스 중에서 Lee, Ryu, Kim의 공통 번역 텍스트인 GE의 2그램 목록을 대표로 제시한다.

표 2. GE ‘as if’ 구문 번역 2그램 목록

Lee			Ryu			Kim		
순위	2그램	빈도	순위	2그램	빈도	순위	2그램	빈도
1	한 것처럼	40	1	한 듯	15	1	것 같았다	63
2	것처럼 말이다	16	2	것 같았다	13	2	한 것처럼	14
3	있는 것처럼	15	3	마치 내가	9	3	되는 듯	10
4	것 같은	14	4	없다는 듯	7	4	게 마치	7
5	마치 내가	13	5	것 같은	6	5	것 같고	6
6	있기라도 한	13	6	한 것처럼	6	6	것처럼 보였다	6
7	되는 것처럼	12	7	되는 양	5			
8	것처럼 보였다	10	8	있기라도 한	5			
9	마치 나를	10	9	있는 것	5			
10	마치 그	9	10	있는 것처럼	5			
11	것 같았다	7						
12	마치 내	7						
13	있는 듯한	6						
14	그 순간	5						
15	마치 자신의	5						
16	한 것	5						

[표 2]에서 Lee, Ryu, Kim 모두 ‘것처럼’, ‘듯/듯한’(의존명사), ‘것 같았다/것 같은/것 같고’(형용사) 등 ‘-처럼’, ‘(-)듯*’, ‘(-)같*’를 주로 사용하는 것을 볼 수 있다. 또한 세 번역가의 2그램 목록에서 공통되는 어휘로 부사 ‘마치’가 있다. Lee는 ‘마치’가 들어가는 2그램이 다섯 개 항목(5번, 9번, 10번, 12번, 15번)이나 될 만큼 ‘마치’를 높은 빈도로 사용하였다. Ryu는 ‘마치 내가’가 빈도 순위 3위 항목이고

5) 여기서 ‘-처럼’은 비교나 비유의 뜻을 나타내는 부사격 조사이다. ‘(-)듯*’는 ‘(-)듯’ + 와일드카드(wildcard) 형태로서, ‘-듯’, ‘-듯이’, ‘-듯한’ 등 어간에 바로 붙는 연결 어미 ‘-듯’과 어미 뒤에 띄어 쓰는 의존명사 ‘듯이’ 등이 쓰인 모든 경우를 포함한다. 참고로 연결 어미로 쓰인 ‘-듯이’의 예문(㉞)과 의존명사 ‘듯이’의 예문(㉟)을 들어보면 각각 다음과 같다. ㉞ 땀이 비 오듯이 쏟아진다. ㉟ 죽은 듯이 누워 있다. ‘(-)같*’의 경우 부사격 조사 ‘같이’와 형용사 ‘같다’(‘같은’ 등 포함)로 쓰인 예를 모두 분석하였다. 조사 ‘같이’가 쓰인 예문(㉡)과 형용사 ‘같다’의 예문(㉢)을 들어보면 각각 다음과 같다. ㉡ 소갈이 일만 하다. ㉢ 마음이 비단 같다/백옥 같은 피부

Kim은 ‘계 마치’를 7회 사용하였다. 세 번역가의 2그램 비교에서 또 다른 눈에 띄는 점은 Lee와 Ryu에 공통된 항목 ‘있기라도 한’이다. Lee는 ‘있기라도 한’을 13회 사용하고 Ryu는 5회 사용하였다. Kim은 ‘as if’ 구문 번역에서 ‘있기라도 한’을 1회 사용하였다. 이렇게 세 번역가 모두가 사용한 보조사 ‘-라도’는 ‘-처럼’ 등의 비유를 나타내는 표현과 결합하여 자주 쓰였다. 이에 따라 ‘as if’ 구문 번역에서 자주 쓰이는 표현 ‘-처럼’, ‘(-)듯*’, ‘(-)같*’와 ‘마치’, ‘-라도’를 중심으로 세 번역가의 번역 문체 차이를 살펴본다. 다음 절에서는 이상 5개 표현을 중심으로 군집 분석 및 t-SNE를 실시한 결과를 기술한다.

3.1. 군집분석 및 t-SNE 시각화 결과

Lee, Ryu, Kim의 11개 TT를 대상으로 수행한 군집분석 결과는 [그림 1]과 같다.

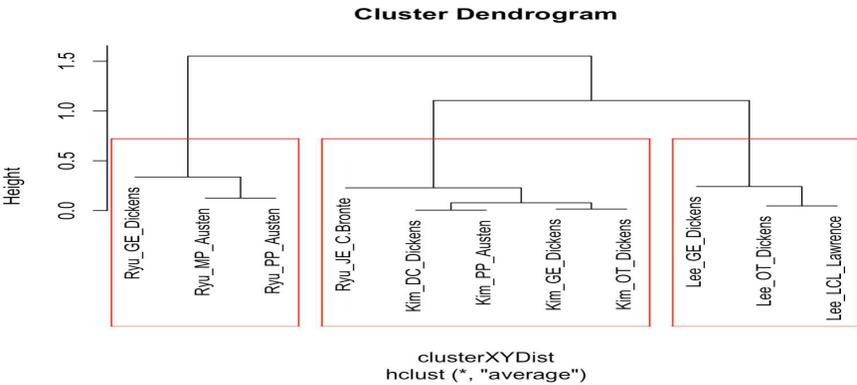


그림 1. 세 번역가 TT 간 계층군집분석 수형도

[그림 1]은 유사도에 따른 세 번역가의 TT 간 분류를 보여준다. 전체 11개의 TT가 총 3개의 그룹으로 군집을 이루는 것을 볼 수 있다. [그림 1]에서 맨 오른쪽 네모 표시를 보면, Lee가 번역한 TT 3개가 모두 하나의 군집을 이룬 것을 볼 수 있다. 이와 달리 Ryu와 Kim은 가운데 네모에서 함께 하나의 군집을 이루고 있다. Ryu가 번역한 ‘Ryu_JE_C. Brontë’가 Kim의 TT들과 같은 군집을 이루고 있는 것이다. 이는 Ryu의 JE가 Kim의 4개 TT와 번역 문체가 유사하다는 것을 나타낸다. 반면 ‘Lee 코퍼스’(GE, OT, LCL)는 ‘Ryu 코퍼스’(GE, MP, PP, JE)와 상대적으로

문체 거리가 멀다고 할 수 있다.

상기 군집분석 결과 확인한 Ryu와 Kim의 TT 간 번역 유사도와 세 번역가 간 차이를 세부적으로 분석하기 위해 t-SNE 차원축소기법을 사용하여 11개의 TT를 2차원 공간에 시각화하였다. 아래 [그림 2]는 Lee, Ryu, Kim의 ‘as if’ 번역코퍼스를 t-SNE 기법으로 분석한 결과이다.

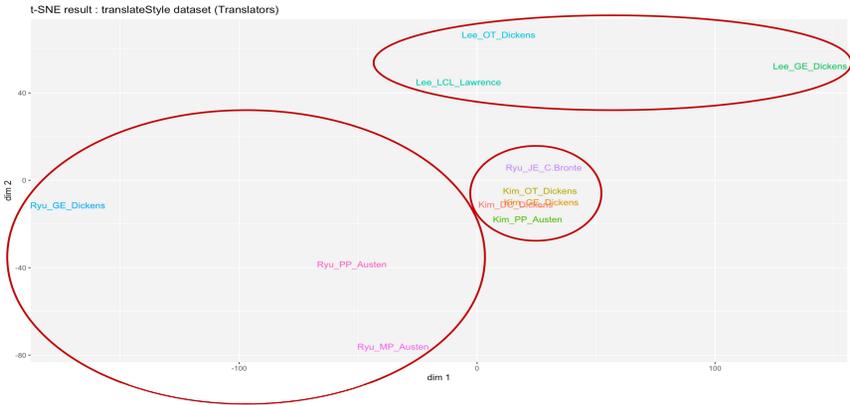


그림 2. t-SNE 시각화 분석 결과

t-SNE는 유사한 데이터들은 가깝게, 유사하지 않은 데이터들은 멀리 떨어지게 나타낸다. 이렇게 데이터 사이의 거리로 데이터들의 유사도를 시각적으로 확인할 수 있다. t-SNE 기법으로 세 번역가가 전체 11개 TT에서 사용한 표현을 분석한 결과 Lee는 t-SNE에서 상단에, Kim은 중간 부분에 위치하고 있으며, Ryu는 중간에서 하단 부분에 넓게 분포하고 있는 것을 볼 수 있다(군집별 동그라미 표시). 또한 중간 부분(제일 작은 동그라미)에 위치한 Kim의 TT들(Kim_OT_Dickens, Kim_GE_Dickens, Kim_DC_Dickens, Kim_PP_Austen)과 ‘Ryu_JE_C. Brontë’가 서로 인접해 있는 것을 볼 수 있다. 이렇게 Ryu 코퍼스(특히 ‘Ryu_JE_C. Brontë’)가 Kim 코퍼스와 가까운 거리에 위치하고 Lee 코퍼스(OT, LCL, GE)가 Ryu 코퍼스(GE, PP, MP, JE) 그리고 Kim 코퍼스(OT, GE, DC, PP)와 상대적으로 떨어져 있는 것을 볼 때, Lee는 ‘as if’ 번역에 있어서 Ryu와, 그리고 Kim과 다른 문체를 형성하고 있다고 할 수 있다. 반면 Ryu 코퍼스(특히 ‘Ryu_JE_C. Brontë’)가 Kim 코퍼스와 가까운 거리에 위치하는 것으로 보아 두 번역가의 ‘as if’ 구문 번역 문체는 비슷하다고 할 수 있다. 다음으로 각 번역가가 번역한 TT들의 군집을 살펴

보면 Kim 코퍼스의 전체 4개 TT가 조밀하게 하나의 군집을 이루고 있고 Lee 코퍼스의 3개 TT도 하나의 군집을 이루고 있는 것을 볼 수 있다. 이에 비해 Ryu 코퍼스는 4개 TT 간 거리가 상대적으로 멀고, 한 TT(Ryu_JE_C. Brontë)는 오히려 다른 번역가 코퍼스(Kim)에 더 가깝게 위치한 것을 확인할 수 있다.

3.2. 카이제곱 및 상관관계 분석 결과

앞 절에서 t-SNE 분석으로 확인한 번역가 간 문체 차이와, 각 번역가코퍼스 내 TT들의 문체를 세부적으로 분석하기 위해 t-SNE 결과에 대한 사후분석으로 카이제곱 검정과 상관관계 분석을 수행하였다. 먼저 Lee, Ryu, Kim의 공통 번역텍스트 GE를 대상으로 세 번역가의 번역 문체를 분석하였다. 이후 Lee, Ryu, Kim 개별 번역가코퍼스 내 전체 TT를 대상으로 한 번역 문체 분석을 진행하였다.

빈도수를 기반으로 ‘-처럼’, ‘(-)같*’, ‘(-)듯*’, ‘마치’, ‘-라도’의 5개 표현을 중심으로 이들 표현의 사용과 두 개 이상의 표현을 조합한 패턴 사용을 분석하였다. 전체 조합 패턴 중에서 사용 빈도 5회 이상인 패턴들을 분석하였다. 분석에 사용된 19개의 패턴은 아래 표와 같다.

표 3. 번역코퍼스에서 5회 이상 사용된 패턴 조합

패턴	빈도	패턴	빈도	패턴	빈도
처럼	147	같	182	듯	237
마치	28	라도	13		
처럼 + 듯	7	마치 + 처럼	81	라도 + 처럼	78
마치 + 듯	70	마치 + 같	62	마치 + 라도	15
라도 + 듯	111	라도 + 같	44	듯 + 같	9
처럼 + 마치 + 듯	6	마치 + 라도 + 처럼	94	마치 + 라도 + 듯	54
마치 + 라도 + 같	25	사용 안 함	165		

3.2.1. 공통텍스트 GE 대상 번역가 간 문체 비교

본 연구에서는 세 번역가의 공통 번역텍스트 GE를 대상으로, ‘as if’ 구문 번역에서 번역가 사이에 문체 차이가 있는지를 확인하기 위해 이원 카이제곱 검정을 수행하였다. 분석 결과, p값이 2.2e-16으로 95% 신뢰구간인 0.05보다 작아 번역가들 사이에 문체 차이가 있다는 결과가 나왔다(X-squared = 403.44, df = 38). 이를

바탕으로 어떤 번역가들 사이에 문체 차이가 나는지 확인하기 위해 상관관계 분석을 추가로 진행하였다. 번역가 간 번역 문체에 상관관계가 있는지 여부를 분석한 결과는 [표 4]와 같다.

표 4. GE 대상 번역가 간 문체 상관관계 분석

번역가	t 값	p 값	상관계수
Lee vs. Ryu	-0.1005	0.921	-0.0236
Lee vs. Kim	-0.4471	0.6601	-0.1048
Kim vs. Ryu	3.3951	0.0032	0.6248

GE ‘as if’ 구문 번역의 번역가 간 문체 상관관계를 분석한 결과, Kim과 Ryu의 경우 p값이 0.0032로 0.05보다 작으므로 상관관계가 있는 것으로 나왔다. 즉, Kim과 Ryu 두 번역가의 GE 번역 문체는 유사하다고 볼 수 있다. 반면 Lee는 각각 Ryu, Kim 두 번역가와 비교 결과에서 p값이 모두 0.05보다 크므로 Lee는 각각 Ryu, Kim 두 번역가와 ‘as if’ 번역 문체가 다르다는 결과가 나왔다.

아래 [그림 3]에는 각 번역가가 GE ‘as if’ 구문을 번역할 때 사용한 표현을 수치화 및 시각화하여 제시하였다. 상기 상관관계 분석 결과에서 더 나아가 각 번역가가 구체적으로 어떤 패턴의 차이를 보이는지 확인할 수 있다. 번역가별 사용 표현(‘-처럼’, ‘(-)듯*’, ‘(-)같*’, ‘마치’, ‘-라도’ 등)을 빈도에 기반하여 막대그래프로 나타내고 번역가별로 사용 빈도 1위에서 3위에 해당하는 항목을 표시하였다.

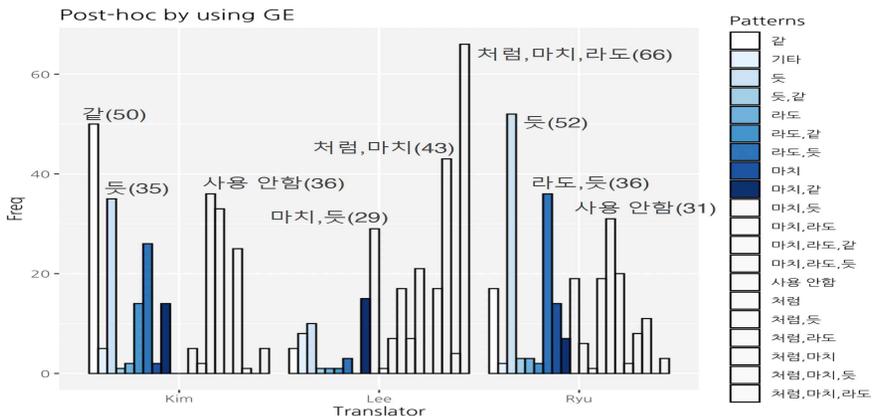


그림 3. 번역가별 GE ‘as if’ 구문 번역 패턴 및 빈도 시각화

빈도 상위 표현들을 중심으로 살펴보면, Kim은 ‘(-)같*’와 ‘사용 안 함’을 많이 사용하였다. ‘사용 안 함’은 ‘-처럼’, ‘(-)같*’, ‘(-)듯*’, ‘마치’, ‘-라도’의 5개 패턴 및 패턴 조합을 사용하지 않은 경우를 말한다. Ryu는 ‘-라도’) + ‘(-)듯*’를 높은 빈도로 사용하였다. Lee는 ‘마치’ + ‘(-)라도’) + ‘-처럼’을 집중적으로 사용하였다. 또한 Kim과 Ryu는 ‘사용 안 함’ 항목의 빈도 순위가 각각 2위와 3위인데, Lee는 ‘사용 안 함’의 순위가 이보다 많이 낮았다. 다시 말해 Lee는 GE에서 상기 패턴 조합을 (하나라도) 사용하여 번역한 경우가 많은 반면, Kim과 Ryu는 이러한 패턴 조합들을 사용하지 않고 번역한 경우가 Lee에 비해 자주 있었다. 이상 GE ‘as if’ 번역의 통계 분석 및 번역가별 번역 패턴 분석에서, ST가 같은데도 세 번역가가 각각 사용한 패턴이 다른 번역가와 구별되는 특징(Saldanha 2011)을 보여 각자 번역가 문체가 존재함을 확인할 수 있었다. Kim과 Ryu는 ‘(-)듯*’와 ‘사용 안 함’의 빈도에서 유사한 패턴을 보여 문체 유사성이 나타난 반면 Lee는 각각 Kim, Ryu와 차이를 보였다.

3.2.2. 개별 번역가코퍼스 내 문체 분석

3.2.2.1. Lee 코퍼스에 나타난 번역가 문체

앞 절의 군집분석 결과에서 하나의 독립된 군집을 형성한 Lee의 TT들을 더 자세히 분석하기 위해 카이제곱 검정과 상관관계 분석을 시행하였다. Lee의 TT들 사이에 번역 문체 차이가 있는지를 카이제곱 검정으로 분석한 결과, p값이 0.05보다 작았다($X^2=97.569, df=36$). 즉, Lee가 번역한 3개 TT 간 ‘as if’ 구문 번역 문체에 차이가 있다는 결과가 나왔다. 앞서 세 번역가의 공통 TT인 GE를 대상으로 카이제곱 검정과 상관관계 분석을 시행한 것처럼 Lee 코퍼스 분석에서도 카이제곱 검정과 상관관계 분석을 시행하였다. Lee 코퍼스의 상관관계 분석 결과는 아래 [표 5]와 같다.

표 5. Lee 코퍼스 TT 간 문체 상관관계 분석

TT	t 값	p 값	상관계수
GE vs. LCL	2.9821	0.008368	0.5860424
GE vs. OT	2.1792	0.04367	0.4672861
LCL vs. OT	3.2459	0.004753	0.6185704

상관관계 분석으로 Lee 코퍼스의 TT별 번역 문체를 비교한 결과 모든 p값이

0.05보다 작아 Lee의 3개 TT의 번역 문체가 유사한 것으로 나왔다. 이렇게 상관관계 분석 결과가 상기 카이제곱 검정 결과와 상반되게 나와, 아래에서 Lee 코퍼스의 번역 패턴 및 빈도를 자세히 살펴본다. [그림 4]는 Lee 코퍼스에서 ‘as if’ 번역에 사용한 패턴을 시각화한 그래프이다.

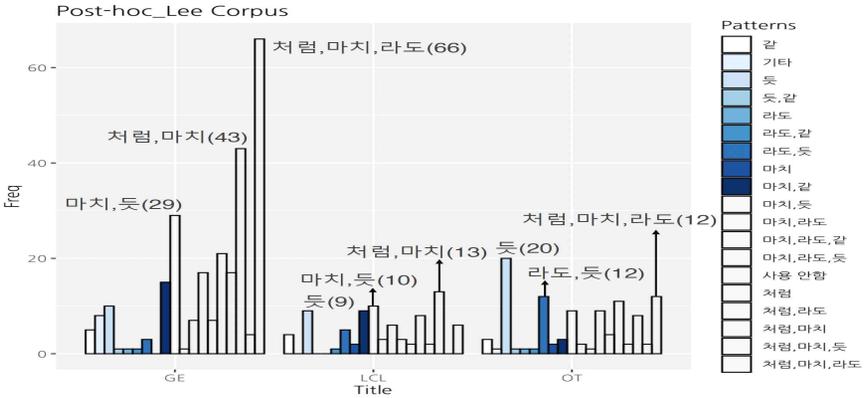


그림 4. Lee 코퍼스 번역 패턴 및 빈도 시각화

빈도 1위에서 3위까지 패턴을 중심으로 보면 LCL과 OT에서는 ‘(-)못*’가 다른 번역 패턴에 비해 사용 빈도가 높았지만, GE에서는 상대적으로 빈도가 높지 않았다. 또한 Lee는 GE에서 ‘마치’ + ‘-처림’을 많이 사용하였으나 OT에서는 ‘(-)못*’을 주로 사용하였다. 실제로 위 상관관계 분석 결과에서 GE와 OT의 상관계수 값은 0.4672861로 Lee의 다른 상관계수 값에 비해 작은 값이 나왔다. 이런 결과를 볼 때 GE에서의 번역 패턴과 OT에서의 번역 패턴 차이로 인해 상기 카이제곱 검정에서 Lee의 TT 간에 문체 차이가 난다는 결과가 나온 것을 알 수 있다.

아래에서는 각 TT에서 예시를 들어 Lee의 번역 패턴을 살펴본다.

[예시 1]

ST: [...] and stood frowning at his boots as if he suspected them of designs against him.

TT-Lee_GE_Dickens: 그러고는 자신의 구두가 그를 해치려는 무슨 음모라도 꾸민다고 의심하는 것처림, 찌푸린 얼굴로 구두를 노려보며 서 있었다.

[예시 1]은 Lee가 ‘as if’ 번역에서 많이 사용하는 패턴 ‘-라도’ + ‘-처림’의 예이

다. Lee는 “사실은 아니지만 그 정도”(홍운혜 2007: 9)라는 의미의 보조사 ‘-라도’를 자주 사용함으로써 영어 ‘as if’의 비유 의미를 강조하고자 한 것으로 보인다.

[예시 2]

ST: [...] the dog running forward, as if conscious that there was no further occasion for his keeping on guard, [...]

TT-Lee_OT_Dickens: 개가 마치 더 이상 감시할 필요가 없는 것을 알아차리기도 한 듯이 앞으로 곧장 달려가 [...]

[예시 2]는 Lee가 ‘마치’ + ‘-라도’ + ‘듯이’를 사용한 예이다. Lee는 모든 TT에서 다른 번역가들보다 자주 그리고 일관되게 부사 ‘마치’를 쓰고 있다.

이상 Lee 코퍼스의 통계 분석 및 번역 패턴 분석에서, ST가 각기 다른 Lee의 3개 TT 모두 공통된 번역 패턴을 보였다는 점에서 번역가코퍼스 내에서 일관된 ‘선택’의 패턴을 보이는 번역가 문체 요건(Saldanha 2011)을 충족한 것을 볼 수 있다. 이상 분석에서 ‘마치’ + (‘-라도’) + ‘-처럼’ 조합을 많이 쓰고 ‘(-)듯*’를 자주 사용하는 반면 ‘(-)같*’는 잘 사용하지 않는 Lee의 번역 패턴을 확인하였다.

3.2.2.2. Ryu 코퍼스에 나타난 번역가 문체

다음은 Ryu 코퍼스 내 4개 TT의 문체 비교를 위해 진행한 카이제곱 검정과 상관관계 분석 결과이다. 카이제곱 검정 결과, p값이 0.05보다 작은 0.004467로 나왔다(X-squared = 85.043, df = 54). 즉, Ryu의 4개 TT에서 ‘as if’ 구문의 번역 문체에 차이가 있는지를 검증한 결과 TT들 사이에 번역 문체 차이가 있다는 결과가 나왔다. 다음으로 TT 간 연관성을 살펴보기 위해 Ryu의 TT별 상관관계 분석을 진행한 결과는 [표 6]과 같다.

표 6. Ryu 코퍼스 TT 간 문체 상관관계 분석

TT	t 값	p 값	상관계수
GE vs. JE	5.3194	5.648e-05	0.7903739
GE vs. MP	5.0651	9.565e-05	0.7755333
GE vs. PP	4.3557	0.0004302	0.7262342
JE vs. MP	6.5633	4.827e-06	0.8467766
JE vs. PP	4.398	0.0003929	0.7295377
MP vs. PP	10.182	1.188e-08	0.9268875

위 표에서 보듯이 모든 TT 간 p값이 0.05보다 작게 나왔다. Ryu 코퍼스 4개 TT의 번역 문체가 유사하다는 결과이다. 지면 제약상 Ryu 코퍼스의 TT별 패턴 및 빈도 그래프는 제시하지 않지만, Ryu의 번역 패턴 분석 결과 다음과 같은 번역가 문체를 확인하였다. Ryu는 모든 TT에서 ‘as if’ 구문 번역에 ‘(-)듯*’를 가장 많이 사용하였고 ‘-처럼’, ‘(-)듯*’, ‘(-)같*’, ‘마치’, ‘-라도’의 5개 단일 패턴 또는 패턴 조합을 사용하지 않는 ‘사용 안 함’의 패턴도 많이 나타났다. 위에서 분석한 Lee와 비교해보면 Ryu는 ‘-처럼’을 잘 사용하지 않는 문체 특징을 보인다. Ryu의 이러한 패턴은 한국어 비번역코퍼스에서 ‘(-)듯*’, ‘-처럼’, ‘(-)같*’ 중 ‘-처럼’이 사용 빈도 1위(연지은 2010: 193-194)라는 점에서 특징적이다. 아래 예시에서 Ryu의 번역 패턴을 살펴본다.

[예시 3]

ST: He had been walking fast about the room, and he stopped, as if suddenly rooted to one spot.

TT-Ryu_JE_Brontë: 그는 빠른 발걸음으로 이리저리 방을 왔다갔다 했다. 그러다 갑자기 한곳에 뿌리라도 내린 듯 멈춰 섰다.

[예시 3]의 같은 문장을 유종호(2004)는 “그러다가 돌연히 한곳에 뿌리가 박혀 버린 것처럼 우뚝 멈춰 섰다.”고 번역했다.

[예시 4]

ST: We found the air as carefully excluded from both [bedrooms], as if air were fatal to life; [...]

TT-Ryu_GE_Dickens: 우리는 공기가 생명에 치명적인 존재라도 되는 듯이 두 방 모두 바깥 공기가 철저히 차단되어 있는 걸 발견했다.

위 예시들에서 보이는 Ryu의 문체 특징은 Ryu가 실체가 아닌 가상 상황을 비유하며 ‘-라도’ + ‘(-)듯*’ 패턴을 자주 사용한다는 점이다. [예시 4]의 같은 문장을 Lee는 “두 방 모두 마치 공기가 생명에 치명적이기라도 한 것처럼”이라고 번역하였다. 아래 [예시 5]에서도 Ryu의 이러한 문체 특징을 확인할 수 있다.

[예시 5]

ST: I could hardly have imagined dear old Joe looking so unlike himself or so like

some extraordinary bird; standing, as he did, speechless, with his tuft of feathers ruffled, and his mouth open as if he wanted a worm.

TT-Ryu_GE_Dickens: 나는 사랑하는 조가 그토록 자기답지 않고, 그토록 특이한 새 같은 모습을 보이는 걸 상상할 수 없었다. 정말이지 그는 벌레라도 잡아먹고 싶다는 듯 깃털을 잔뜩 곤두세운 새 같은 모습으로, [...]

이처럼 Ryu는 가상 상황 비유에 주로 쓰이는 직유 표현 ‘-처럼’ 대신, 짐작이나 추측의 의미로 쓰이는 의존명사 ‘듯(이)’⁶⁾를 잘 사용한다는 점이 특징적이다. Ryu의 이 같은 패턴은 다른 번역가들이 (가상) 비유 의미를 표현할 때 주로 ‘-처럼’을 쓰는 것과 대비된다.

3.2.2.3. Kim 코퍼스에 나타난 번역가 문체

마지막으로 Kim의 TT들에 대한 카이제곱 검정을 진행하였다. 분석 결과, p값이 0.05보다 커($X^2 = 49.661$, $df = 51$) Kim이 번역한 4개 TT 사이에 ‘as if’ 번역 문체에 차이가 없다는 결과가 나왔다. 이후 상관관계 분석을 진행하여 Kim의 TT 문체를 심층 비교하였다. 분석 결과는 아래 표와 같다.

표 7. Kim 코퍼스 TT 간 문체 상관관계 분석

TT	t 값	p 값	상관계수
DC vs. GE	12.151	1.715e-09	0.9498607
DC vs. OT	14.153	1.82e-10	0.9623073
DC vs. PP	8.3291	3.274e07	0.901437
GE vs. OT	10.95	7.662e-09	0.9392904
GE vs. PP	5.9536	2.022e-05	0.8300552
OT vs. PP	5.8117	2.65e-05	0.8237452

Kim 코퍼스 내 TT 간 번역 문체를 상관관계 분석으로 비교한 결과, 모든 TT 간 p값이 0.05보다 작게 나왔다. 이는 Kim의 번역 문체가 전체 4개 TT에서 모두 유사하다는 의미이다. 앞서 제시한 Lee의 TT 간 문체 상관관계 분석 결과([표 5])와 비교할 때, [표 7]의 Kim의 상관계수 값이 전체적으로 매우 큰 것을 확인할 수

6) 연결 어미로 쓰인 ‘-듯’, ‘-듯이’와 의존명사 ‘듯’, ‘듯이’는 띄어쓰기로 구별할 수 있다. ‘돈을 물 쓰듯이 쓴다.’에서 ‘-듯이’는 비슷한 사실을 비교할 때 쓰는 연결 어미로, 어간에 붙여 쓴다. 반면 짐작이나 추측의 의미를 나타내는 의존명사 ‘듯이’는 ‘죽은 듯이 누워 있다.’처럼 어미 ‘-은/-는/-을’ 뒤에 띄어 쓴다.

있다. 이는 Kim의 경우 번역 내용이 달라도, 즉 각기 다른 ST를 번역해도 Kim의 TT들에서는 ‘as if’ 번역 문체에 차이가 나지 않는다는 것을 의미한다. 앞서 [그림 2] t-SNE 시각화에서 Ryu나 Lee의 TT들은 서로 상대적으로 먼 거리에 위치한 반면 Kim의 TT 4개는 유사도가 높아 거의 중첩되게 조밀한 군집을 형성한 것과 일치하는 결과이다.

아래 [그림 5]에서 Kim 코퍼스의 TT 그래프 모두가 유사한 패턴을 보이는 것을 확인할 수 있다.

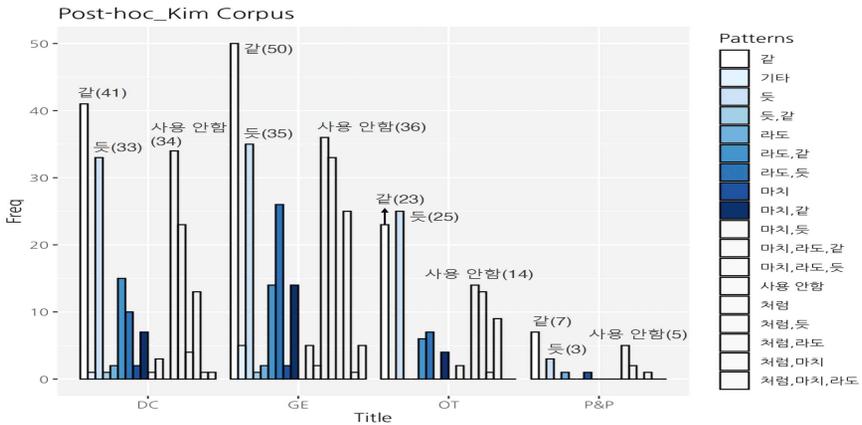


그림 5. Kim 코퍼스 번역 패턴 및 빈도 시각화

번역 패턴 분석 결과, Kim 코퍼스에서 ‘as if’ 번역에 가장 많이 사용한 패턴은 단일 패턴 ‘(-)갈*’과 ‘(-)똥*’이다. 그리고 ‘-처럼’, ‘(-)똥*’, ‘(-)갈*’, ‘마치’, ‘-라도’의 5개 패턴 중 아무것도 사용하지 않은 ‘사용 안 함’도 높은 빈도로 나타났다. Kim은 ‘(-)갈*’이나 ‘(-)똥*’ 하나만은 많이 사용하나 ‘마치’나 ‘-라도’와 함께는 잘 쓰지 않았다. 그리고 ‘-처럼’을 잘 사용하지 않는 문체 특징을 보였다. 이러한 특징은 위에서 분석한 Ryu와 유사하다. 앞서 공통텍스트 GE를 대상으로 한 번역기간 상관관계 분석에서 Kim과 Ryu의 번역 문체가 유사한 것으로 나온 것([표 4] 참조)은 이러한 패턴의 일치에서 기인한 것으로 보인다. 반면 Kim은 Lee의 빈도 상위 패턴인 ‘마치’ + ‘(-라도)’ + ‘-처럼’을 거의 쓰지 않는다는 점에서 Lee와 차이가 난다.

아래 예시들에서는 Kim의 ‘것 같았다’ 사용 패턴을 볼 수 있다.

[예시 6]

ST: The man was limping on towards this latter, as if he were the pirate come to life, and come down, and going back to hook himself up again.

TT-Kim_GE_Dickens: 사내는 교수대를 향해 절뚝거리면서 나아가는데, 해적이 살아서 내려왔다가 다시 목을 매달려고 돌아가는 것 같았다.

상기 예시들에서 Ryu가 원문의 ‘as if’ 비교·비유 의미를 짐작·추측의 의미인 ‘듯’을 사용해 번역한 것처럼 [예시 6]에서 Kim은 ‘것 같았다’를 사용해 번역한다.

[예시 7]

ST: [...] but, the old gentleman, eyeing him with an expression of dislike, look anxiously round, as if he contemplated running away himself: [...]

TT-Kim_OT_Dickens: 노신사는 경멸스런 표정으로 상대를 바라보곤 걱정스러운 눈길로 주변을 둘러보는 자세가 이번엔 자신이 도망칠 기회만 엿보는 것 같았다.

[예시 8]

ST: With that he looked back, and nodded at his dead plant, and then cast his eyes about him in walking out of the yard, as if he were considering what other pot would go best in his place.

TT-Kim_GE_Dickens: 죽은 꽃에게 고개를 끄덕인 다음, 마당을 이리저리 살피면서 걷는 모습이 빈자리에 어떤 화분을 갖다놓는 게 제일 좋을까 헤아리는 것 같았다.

[예시 7]에서 Kim은 인물의 행동에 근거한 추측이나 판단의 의미로 ‘것 같았다’를 사용하고 [예시 8]에서는 “as if he were” 가정법을 활용한 비유(사형수를 ‘죽은 꽃’이나 ‘화분’에 비유)에 ‘것 같았다’를 사용하였다. 상기한 [예시 6]의 “as if he were” 비유에서도 ‘것 같았다’를 사용하였다. 이처럼 Kim은 ‘~ 자세가’나 ‘~ 모습이’, ‘-ㄴ 걸 보면’ 등의 표현으로 인물의 행동을 묘사한 후 ‘-ㄴ 것 같았다’로 문장을 끝맺는 패턴을 보인다. 이렇게 ‘~ 자세가’나 ‘~ 모습이’ 등이 나오고 뒤에 ‘것 같았다’가 붙는 문장 형태가 반복적으로 나오면서 ‘~ 자세가’나 ‘~ 모습이’가 추측이나 판단의 근거로 제시된 것 같은 느낌을 주기 때문에 ‘것 같았다’를 쓰면 [예시 8]과 같은 독특한 비유가 효과적으로 살지 않는다.) ‘자세가/모습이/-ㄴ 걸 보면/-

ㄴ 게 + 것 같았다'의 형태를 'as if' 비유를 번역할 때 사용하는 것의 문제는 Kim이 대체로 '것 같았다'나 '것처럼 보였다' 등으로 번역되는 '(It) seemed'를 번역할 때도 동일하게 '것 같았다'를 사용한다는 점에서도 알 수 있다. 지면 제약상 예문을 제시하지는 않지만, Kim은 GE를 번역하며 주로 비유 의미를 나타내는 'as if'와 짐작이나 추측을 나타내는 'seemed'를 달리 번역하지 않고 동일하게 '것 같았다'를 사용한 경우가 많았다.⁸⁾

상기 분석에서 Lee, Ryu, Kim 모두 'as if' 구문 번역에 있어서 다른 번역가와 구별되는, 일관된 패턴을 보여 각자 번역가 문체를 형성하고 있는 것을 확인하였다. 또한 군집분석과 t-SNE 분석으로 Kim과 Ryu 사이의 문체 거리가 가까움을 시각적으로 확인한 후, 카이제곱 및 상관관계 분석을 추가 진행하여 Kim과 Ryu의 유사한 문체 특징을 발견하였다. Kim과 Ryu의 각 TT를 분석하여 '(-)같*', '(-)듯*'의 사용 패턴과 '사용 안 함'의 고빈도 그리고 '-처럼'의 저빈도에서 두 번역가의 문체 유사성이 나타남을 구체적으로 밝혔다. 반면 군집분석과 t-SNE 시각화 분석에서 각각 Ryu, Kim과 거리가 떨어져 독립적인 군집을 형성한 Lee는 TT에서 '마치' + '-처럼'의 고빈도 사용 및 '(-)같*'와 '사용 안 함'의 저빈도 사용 패턴을 보여 'as if' 구문 번역에서 Ryu, Kim과 확연한 문체 차이를 보였다.

Kim은 세 번역가 중 유일하게 카이제곱 검정과 상관관계 분석 모두에서 자신의 TT들의 'as if' 번역 문체에 차이가 없다는 결과가 나왔다. 각기 다른 ST를 번역한 Kim의 4개 TT는 ST 내용과 상관없이 균일한 문체를 보였다. 'as if' 구문 번역과 관련해서는 ST 내용이 TT 문체에 끼치는 영향보다 Kim의 번역가 문체가 끼치는 영향이 더 크다고 볼 수 있는 것이다.

앞서 제시한 번역가별 TT 간 문체 상관관계 분석 결과를 비교·종합하여, 세 번역가가 ST의 영향을 받는 정도도 파악할 수 있었다. 번역가코퍼스별 상관계수를 비교한 결과, 세 번역가의 'as if' 번역 문체에 미치는 ST의 영향은 차이가 있었다. Kim은 각 TT 간 상관계수가 고르게 크게 나와(0.8237-0.9623) ST 영향을 거의 받지 않는다고 할 수 있는 반면, Ryu는 동일 저자의 다른 두 작품(PP_Austen, MP_Austen)을 제외하고는 Kim에 비해 상대적으로 TT 간 상관계수 값이 작아(0.7262-0.8468) Kim보다 Ryu가 ST의 영향을 더 받는다고 할 수 있다. Lee는 TT 간 상관계수 값이 전반적으로 세 번역가 중 가장 작고(0.4673-0.6186), 동일 저자

7) '-은 것 같다', '-은 듯 하다' 등은 한국어의 대표적인 추측표현이다(김동욱 2000).

8) 'seemed' 번역 예시 및 관련 논의는 신혜정(2019) 참조

디킨스의 두 작품(GE_Dickens, OT_Dickens) 간 상관관계수 값도 가장 작아(0.4673) Kim과 Ryu에 비해 ST의 내용에 따라 ‘as if’ 번역 문체가 달라지는 정도가 큰 것을 알 수 있다. 본 연구는 TT의 일부인 ‘as if’ 번역문으로 범위를 한정하여 분석하였으므로, 위와 같은 번역가 3인의 문체에 대한 ST의 영향 정도가 Lee, Ryu, Kim 각 번역가의 TT 전체(full text)나 번역가코퍼스 전체에서도 동일하게 나타나리라고 일반화할 수는 없다. 이를 밝히기 위해서는 분석 범위와 분석 대상을 확대한 추가 연구가 필요하다.

4. 결론

본 연구에서 확인된 세 번역가의 ‘as if’ 번역 문체는 다음과 같다. Lee는 뒤에 이어지는 내용이 무언가를 빗대어 표현하는 비유임을 직접적으로 표시하는 부사 ‘마치’와 직유 표시어 ‘-처럼’의 패턴 조합을 사용하는 고유한 ‘as if’ 번역 문체를 보였다. 즉, ‘마치’ + (‘-라도’) + ‘-처럼’의 패턴을 일관되게 사용하여 원문의 ‘as if’ 비유 의미가 확실하게 드러나도록 표현하였다. 이는 비유 의미를 드러내는 ‘마치’ 대신 짐작이나 추측의 의미로 쓰이는 ‘듯(이)’나 ‘것 같았다’를 주로 사용한 Ryu나 Kim과 구별되는 특징이다. Ryu는 ‘사실은 아니지만 그 정도’라는 비유 의미가 내포된 보조사 ‘-라도’에 짐작이나 추측을 나타내는 의존명사 ‘듯(이)’를 결합하여 쓰는 특이한 패턴을 보였다. Lee도 Ryu와 같이 가상 상황 비유에 ‘-라도’를 많이 사용하였지만, Lee는 ‘-라도’에 직유 표현 ‘-처럼’을 결합하여 사용하였다는 점이 다르다. Kim은 ‘-니 걸 보면/-니 게/자세가/모습이 + ... 것 같았다’와 같은 독특한 ‘as if’ 번역 패턴을 보였다. ‘as if’ 비유 번역에 ‘것 같았다’라는 추측의 표현을 써서 문장을 끝맺는 Kim의 번역 패턴은 특히 한국어 직유 패턴에서 크게 벗어나는 아웃라이어(outlier) 같은 존재라는 점에서 Kim의 ‘as if’ 번역 문체 표지이며 그의 문체 지문(stylistic fingerprint)이다.

이와 같은 번역가별 문체 특징에 워드스미스의 부정 핵심어(negative keyword) 개념이라 할 수 있는, ‘기대치에 비해 확연히 차이 나게 사용 빈도가 낮은 표현’까지 분석한 결과를 추가해 각 번역가의 번역 패턴들이 종합, 축적되면 개별 번역가의 문체 프로파일링(profiling)도 가능하다. 본 연구에서 확인한 번역가별 문체 정량화 결과는 일정 정도 쌓여 데이터베이스화되면 추후에 번역가 미상 작품의 번

역가 판별(translation attribution)에 쓰일 수 있으며 번역본의 표절 여부를 확인 (Caballero et al. 2021)하는 데도 활용할 수 있다.

텍스트 마이닝 기법 및 다양한 통계 분석 방법을 적용하여 번역가 문체를 계량적으로 분석한 본 연구는 각 번역가의 번역 문체를 확인하고 ‘눈에 보이지 않는’ 번역가 간 문체 거리를 시각화하여 나타내었다. 이러한 정량 연구 결과를 바탕으로 텍스트들을 비교하고 세부 분석하여 번역가 간 문체 유사점과 차이점을 밝힘으로써 문체 연구에 유용한 정량분석 방법을 제시하였다. 탐구적 통계 분석 방법을 적용한 본 연구 결과는 후속 연구를 이끌어 기존의 직관에 의존한 정성적인 문체 연구 성과를 더 풍성하게 할 수 있다.

본 연구는 그간 영어 번역(영역)에 비해 연구가 미흡했던 한국어 번역(한역) 텍스트의 번역가 문체판별에 군집분석, t-SNE, 카이제곱 검정 등 계량적 기법을 성공적으로 적용할 수 있음을 보였다. 특히 문학 텍스트에서 한국어의 (보)조사, 어미, 의존명사 등의 기능어가 효과적인 번역가 문체 변별자질임을 보였다는 점에 의의가 있다. 본 연구에서 분석한 ‘as if’ 번역 패턴, 다시 말해 ‘(-)같*’, ‘-처럼’, ‘(-)듯*’, ‘마치’, ‘-라도’의 사용 패턴과 ‘사용 안 함’의 패턴은 ST 독립적으로 (ST-independent), 즉 ST가 달라도 적용 가능한 ‘번역가 문체판별요인’으로서 향후 한국어 텍스트의 번역(가) 문체 연구에 활용할 수 있다.

‘번역 텍스트의 문체’가 아닌 ‘번역가의 문체’를 규명한 본 연구는 동일 ST를 번역한 복수 TT 코퍼스는 물론 이들 TT를 번역한 동일 번역가들의 개별 번역가 코퍼스 분석을 함께 진행해야 하는 제한으로 인해 분석 데이터 및 분석 범위가 제한될 수밖에 없는 한계가 있다. 후속 연구에서는 이러한 제약을 줄일 수 있도록 다른 자질들로 범위를 확대하거나 전체 텍스트를 대상으로 번역 패턴을 분석하여 번역가 문체를 연구하는 시각을 넓히고자 한다.

참고문헌

- 김동욱. (2000). 「한국어 추측표현의 의미차이에 관한 연구 - ‘ㄴ 것 같다’, ‘ㄴ 듯 하다’와 ‘ㄴ가 보다’, ‘ㄴ 모양이다’의 의미차이를 중심으로. 『국어학』 35: 171-197.
- 김일환·이도길. (2015). 「저자 판별을 위한 전산 문체론 - 초기 현대소설을 대상으로. 『국어국문학』 170: 207-239.
- 신혜정. (2019). 코퍼스 분석을 활용한 ‘as if’ 비유 번역 전략 분석 - 찰스 디킨즈(Charles Dickens)의 *Great Expectations* 한역본을 중심으로. 한국외국어대학교 통번역대학원 박사학위논문.
- 신혜정. (2021). 「번역가 문체의 코퍼스 주도적 연구 -‘as if’ 비유 번역을 중심으로-. 『통역과 번역』 23(2): 77-104.
- 연지은. (2010). 「한국어 직유 표현의 학습 단계별 위계화. 『인문논총』 19: 188-215.
- 이창수. (2015). 「컴퓨터 문체 분석법을 활용한 번역가 문체 연구. 『통역과 번역』 17(3): 119-137.
- 이창수. (2020). 「번역문체 분석 방법으로서의 주성분분석(PCA). 『통역과 번역』 22(3): 117-139.
- 최희경. (2020). 「한국문학 번역에서 고찰하는 번역가 문체: 세 번역가에 대한 코퍼스 주도적 분석. 『인문사회21』 11(6): 1325-1340.
- 하길중. (2005). 「중·근대 국어의 직유 표현. 『한글』 268: 61-85.
- 홍윤혜. (2007). 「보조사 ‘-(이)라도’의 의미와 분포적 특성: ‘-(이)나’와 ‘(이)나마’와의 비교를 중심으로. 『한국어교육』 18(2): 487-506.
- Baker, M. (2000). Towards a methodology for investigating the style of a literary translator. *Target* 12: 241-266.
- Caballero, C., Calvo, H. and Batyrshin, I. (2021). On explainable features for translatorship attribution: Unveiling the translator's style with causality. *IEEE Access* 9: 93195-93208.
- Hotho, A., Nürnberger, A. and Paaß, G. (2005). A brief survey of text mining. *GLDV Journal for Computational Linguistics and Language Technology* 20(1): 19-62.
- Leech, G. and Short, M. (2007). *Style in Fiction: A Linguistic Introduction to English Fictional Prose* (2nd edn.). Harlow: Pearson Longman.
- Mahlberg, M. (2007). A corpus stylistic perspective on Dickens' *Great Expectations*. In Lambrou, M. and P. Stockwell (eds.), *Contemporary Stylistics*. London/New York: Continuum, 19-31.
- Mahlberg, M. (2013). *Corpus Stylistics and Dickens's Fiction*. London/New York: Routledge.
- Munday, J. (2007). *Style and Ideology in Translation: Latin American Writing in English*. New York: Routledge.
- Saldanha, G. (2011). Translator style: Methodological considerations. *The Translator* 17: 25-50.
- Van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research* 9: 2579-2605.

[분석텍스트]

- Austen, J. (1813). *Pride and Prejudice*. Project Gutenberg.
- 류경희 역. (2017). 『오만과 편견』. 파주: 문학동네.
- 김옥수 역. (2018). 『오만과 편견』. 서울: 비꽃.

Austen, J. (1814). *Mansfield Park*. Project Gutenberg.

류경희 역. (2016). 『맨스필드 파크』. 서울: 시공사.

Brontë, C. (1847). *Jane Eyre*. Project Gutenberg.

유종호 역. (2004). 『제인 에어 1, 2』. 서울: 민음사.

류경희 역. (2010). 『제인 에어 1, 2』. 서울: 웅진씽크빅.

Dickens, C. (1837). *Oliver Twist*. Project Gutenberg.

김옥수 역. (2016). 『올리버 트위스트 1, 2』. 서울: 비꽃.

이인규 역. (2018). 『올리버 트위스트 1, 2』. 서울: 민음사.

Dickens, C. (1849). *David Copperfield*. Project Gutenberg.

김옥수 역. (2018). 『데이비드 코퍼필드 1, 2, 3』. 서울: 비꽃.

Dickens, C. (1861). *Great Expectations*. Project Gutenberg.

이인규 역. (2009). 『위대한 유산 1, 2』. 서울: 민음사.

류경희 역. (2014). 『위대한 유산 상, 하』. 파주: 열린책들.

김옥수 역. (2016). 『위대한 유산 1, 2』. 서울: 비꽃.

Lawrence, D. H. (1928). *Lady Chatterley's Lover*. Project Gutenberg.

이인규 역. (2003). 『채털리 부인의 연인 1, 2』. 서울: 민음사.

This paper was received on 1 April 2022; revised on 10 June 2022; and accepted on 20 June 2022.

Author's email address

hjshin@chosun.ac.kr

simon.seongmin.mun@gmail.com

About the author

Hyejung Shin is an Associate Professor at the Department of English Language and Literature, Chosun University, South Korea. Her research interests center on literary translation, corpus stylistics, and translation teaching.

Seongmin Mun is a postdoctoral researcher at Chosun University. His research interests include natural language processing, machine learning, and data visualization. He is currently working as a postdoctoral researcher and a lecturer for computational linguistics.