

번역 연구에서 코퍼스의 본질: 정량과 정성*

조 준 형
(고려대)

1. 서론

번역이 체계화된 학문으로서 인문학 연구의 주제로 고려되기 시작한 것은 매우 최근의 일이다. 그렇지만 현대 번역학에서 논의되는 번역에 대한 수많은 담론의 핵심적인 주제는 그 이전부터 이미 제기되어 왔다고 할 수 있다. 번역은 서로 다른 언어 공동체 간의 접촉 과정에서 자연스럽게 생겨난 오래된 인간 활동이라고 할 수 있다. 고대 시대부터 그리스어로 쓰인 작품이 라틴어로 번역되었고, 라틴어 서적들이 유럽의 여러 민중어로 옮겨졌으며, 혹은 이슬람의 발달된 문명이 번역을 통해 유럽 대륙으로 전파되었다. 여기에는 언어적, 정치적, 경제적, 사회적인 목적이 복합적으로 작용했다고 할 수 있다. 그리고 그 과정에서 수많은 실제 번역가들이 생각한 번역에 대한 담론들이 바로 현대 번역학의 주요 논제로 이어지고 있다. 현대 번역학의 범주에 실제와 이론이 교차하는 것

* 본 연구는 2015년 한국번역학회 봄 학술대회 발표를 수정·보완한 것임.

은 바로 번역의 이러한 태생적 성격에서 비롯되었다고 할 수 있다.

따라서 코퍼스라는 도구가 실제적인 차원에서 번역 연구에 활용되는 것은 지극히 자연스러운 현상이라고 말할 수도 있다. 언어학에서 코퍼스가 본격적으로 활용된 것은 컴퓨터의 발달이 원인이기도 했지만, 언어 현상을 설명하고자 하는 이론들의 실질적인 보편화를 위해서 가공된 자료가 아닌 실제 사례들을 관찰하는 경험적 연구의 중요성이 대두되었기 때문이다. 이러한 경향은 번역학으로 이어져 태생적으로 경험적일 수밖에 없는 번역 연구에서 코퍼스를 활용하는 것은 자연스러운 현상이라고 할 수 있다.

번역 연구에서 코퍼스를 활용한다는 것은 경험적 태도로서 번역의 실제 양태를 살펴보기 위해서이다. 다시 말해서 실제 번역 텍스트에서 번역 사례를 선별해서 이를 관찰하고 최종적으로 번역 현상을 보편적인 이론으로 설명하기 위해서이다. 그럼에도 불구하고 코퍼스 관련 번역 연구가 항상 긍정적인 부분만을 가지고 있는 것은 아니다. 현대 코퍼스 기반 번역학에서 사용하는 여러 기법들은 기존의 코퍼스 언어학에서 제안된 기법들이 수용되어 활용되는 것이 일반적인데, 그 기법들이 너무 전문화되다보니 번역 연구가 번역에 대한 본질이 아닌 형식에 치중되는 경향을 보이기도 한다. 코퍼스를 기반으로 한 번역 연구에서 가장 중요한 것은 번역이라는 개념을 이해하는 데 단초를 제공하는 질적 자료들, 즉 코퍼스에 포함된 번역 사례들이다. 그런데 코퍼스 기반 연구가 사례들을 기술하면서 질적인 특성(정성)과 계량적인 특성(정량)을 복합적으로 활용하는데, 코퍼스 연구의 전문화, 대용량 코퍼스 처리 그리고 자동번역과 같은 응용 연구들과 연계되면서 정량적 기술 방식에 치중된 태도가 많이 나타나고 있는 것이 사실이다. 물론 번역 연구에서 정량적 분석이 잘못되었다는 것은 아니다. 정량적 자료들이 정성적 자료를 해석하는 데 도움이 되는 것은 사실이며, 자동번역에서는 정량적 정보들은 매우 중요한 자료가 되기 때문이다. 그렇지만 정량적 정보들을 활용하면서 비판적 시각 없이 이를 그대로 수용할 경우에는 여러 가지 문제가 나타날 수 있다. 정량적 정보들이 신뢰성을 가지기 위해서는 명확한 기준에 의해 이 자료들을 엄밀하게 해석하려는 태도가 필요하다.

이러한 관점에서 본고에서는 코퍼스 기반 연구에서 정성적 태도와 정량적 태도의 본질을 살펴보고, 코퍼스 기반 연구에서 가장 핵심이 되는 본질이 무엇인지 알아보려고 한다. 다시 말해서 어떤 비판적 태도 없이 정량적 분석만으로

코퍼스에 의한 번역 연구를 한다는 것이 어떤 문제점을 가질 수 있는지를 살펴 보려는 것이다. 그래서 본 연구는 기존 연구자들보다는 번역 연구에서 코퍼스를 활용하고자 하는 신진 연구자들에 대한 일종의 조언이며, 정량적 분석 태도에 의한 코퍼스 기반 번역 연구에 대한 비판적 성찰이라고 할 수 있다.

2. 코퍼스 기반 번역 연구

번역 연구에서 ‘코퍼스’를 활용한다는 것은 어떤 의미를 가지는가? 코퍼스 기반 번역 연구의 본질에 대한 이 물음은 넓은 의미에서 언어 연구에서 코퍼스의 개념이 무엇인지에 대한 해답을 찾아가는 것이라고도 할 수 있다. 기본적으로 번역 코퍼스는 ‘번역 텍스트의 집합체’이다. 그렇지만 이 집합체는 단순히 텍스트의 물리적 저장고가 아니라, 풍부한 번역의 사례들을 담고 있는 번역 데이터베이스이자, 번역 표본들이라고 할 수 있다. 그래서 잘 구성된 번역 코퍼스는 번역 사례를 관찰하기 위한 가장 훌륭한 원천 자료가 되는 것이다. 이러한 관점에서 코퍼스는 번역 연구의 출발점이자 도착점이라고 말할 수도 있을 것이다.¹⁾

그런데 코퍼스를 처리하는 기법이 전문화될수록, 대용량 코퍼스가 요구될수록 그리고 번역 연구의 또 다른 범주인 자동번역과 연계가 되면서 번역의 질적 이해(정성 분석)보다는, 번역을 하나의 데이터로 인식하여 계량적·확률적으로 처리(정량 분석)하려는 경향이 많이 나타나고 있다. 물론 정량 분석이 결코 잘못된 것은 아니지만 번역 현상의 인문학적 설명이라는 본질은 외면당한 채 번역의 형식화에만 치중된 분석은 분명 비판적인 태도로 경계를 해야 할 부분이라고 할 수 있다. 정량적 자료들에 대한 맹신이 자칫 번역 현상의 잘못된 일반화로 이어질 수 있기 때문이다.

1) 코퍼스의 인식론적 개념 및 코퍼스 번역학에 대한 국내 연구로는 정호정(2003), 조준형(2012), 조준형(2014)를 참조할 수 있다.

2.1. 번역 연구에서 코퍼스의 역할

‘번역학’이란 용어가 탄생한 것은 1970년대 초이며²⁾, 현대 번역학의 주요 논저들이 나타나기 시작한 것이 20세기 중반이라고 한다면, 번역에 관한 담론은 이 보다 훨씬 앞선 고대시대부터 시작되었다고 할 수 있다. 그리스의 정치가이자 웅변가인 키케로(Marcus Tullius Cicero)는 기원전에 이미 번역의 문제를 언급하고 있었다(Ballard 1992 : 39-42). 마찬가지로 로마시대에 중세 그리스어 성경을 라틴어로 번역한 성 히에로니무스(Saint Jerome)는 자신의 번역 경험을 바탕으로 번역에 대한 확고한 개념을 정립하고 있었다.³⁾ 번역사를 보면 각 시대별로 번역에 대한 다양한 개념과 주장이 있었음을 확인할 수 있다.⁴⁾ 그리고 긴 역사의 흐름 속에서 나온 번역에 관한 담론들은 직역/의역의 대립, 현지화/이국화, 번역불가능성처럼 지금도 여전히 중요한 번역의 주제가 되는 것들이었다. 그런데 이러한 번역에 대한 담론들은 무에서 창조된 ‘선형적’인 담론들이 아니라, 각 시기별로 실제 번역을 했던 번역가들의 ‘경험의 산물’이었다고 할 수 있다. 키케로와 성 히에로니무스 모두 그 자신이 번역을 하였으며, 이를 통해 자신들만의 번역에 관한 정의를 제시한 것이라고 할 수 있다.

조르주 무냉(George Mounin)의 『불충실한 미녀(Belles Infidèles)』(1994)가 과학적 엄밀성을 가진 번역불가능성에 관한 논의라면 실제 번역불가능성에 관한 담론이 제기되었던 17세기에는 뒤벨레 등과 같은 문인들이 당시의 문화어였던 라틴어에 대한 민중어 프랑스어의 위상을 높이기 위한 필연적 담론이었다(조재룡 2009: 157-158 참조). 과거의 시기를 넘어서 현대에서도 현대 통번역 이론에서 자주 언급되는 파리통번역대학의 해석이론도 실제 통역가인 셀레스코

2) 번역학을 의미하는 영어 *translation studies*는 홈즈(Holmes)의 1972년 *The Name and Nature of Translation Studies*에서 처음으로 등장하며, 프랑스어 *traductologie*는 같은 시기에 캐나다 번역학자인 해리스(Brian Harris)에 의해 제안되었다(이영훈 2013: 193-197 참조). 반면 한국어 명칭인 ‘번역학’은 이영훈(2013: 209)에 따르면, 1984년 김효중의 논문에서 처음으로 나타난다.

3) 키케로와 성 히에로니무스의 번역론에 대해서는 조재룡(2009: 145-148), 김화영(1999) 참조.

4) 번역사에 대한 자세한 내용은 발라르(Ballard 1992), 베이커(Baker(ed) 1998), 조재룡(2009), 김화영(1999) 참조.

비치(Danica Seleskovitch)의 통역 경험에 의해 태어났으며, 캐나다 번역이론가인 비네와 다르벨네(Vinay & Darbelnet)의 그 유명한 『불어와 영어의 비교문체론(*Stylistique comparée du français et de l'anglais*)』도 영어-프랑스어 번역 텍스트로 구성된 방대한 자료의 관찰에 의한 경험 연구의 결과라고 할 수 있다.⁵⁾

이렇듯 여러 학자들 혹은 실무자들이 제안한 번역에 대한 사유는 번역 경험에 의해 혹은 번역 자료의 관찰에서 출발하고 있음을 알 수 있다.⁶⁾ 경험적 관찰, 다시 말해서 번역 텍스트의 관찰은 곧 ‘코퍼스’라는 기존 언어학에서 경험적 자료로서 역할을 하고 있었던 도구의 필요성을 부각시켰다. 번역 연구에서 경험 연구는 “맥락으로부터 독립적인 법칙 또는 이론의 형성에만 있는 것이 아니라 복잡하고 다면적인 성질을 가진 번역을 이해하고 번역문제 해결을 위한 지식을 만드는 일”(강지혜 2013: 23)이며, 경험 연구의 한 방법인 사례 연구는 “조사 대상이 되는 특정한 단위를 맥락 속에서 세밀하게 조사하는 방법”(Ibid. 17)이다. 그렇기 때문에 풍부한 번역 사례를 가진 코퍼스는 최소한 번역 현상 자체를 연구하는 경우에 중요한 역할을 할 수 있다. 번역이라는 것은 단순히 형식적인 언어의 이동이 아니라, 의미 단위 혹은 사고 단위의 이동이다. 그리고 이 이동은 단순한 원리에 의해서가 아니라 다양한 매개변수(예를 들어, 언어, 문화, 번역가, 독자 등)에 의해 매우 복잡한 경로를 따르게 된다. 이러한 이유로 번역 연구는 수학적 함수의 개념이 아닌, 인문학적 사유가 될 수밖에 없다.⁷⁾ 그래서 현대 번역학에서 번역의 원리를 설명하기 위해서 다양한 번역 사례의 관

-
- 5) 이 저작은 실제 코퍼스를 활용한 연구의 대표적인 예라고 할 수 있으며, 이후 ‘비교문체론’이라는 이름이 들어간 많은 번역 연구의 출발점이 되었다. 비네와 다르벨네는 『불어와 영어의 비교문체론』을 작성하기 위해 코퍼스로서 캐나다 국회의 의사회의를 토대로 구축된 *Hansard Corpus*를 참조하였다.
- 6) 모든 번역 연구가 경험적 연구에서 출발한다고 말할 수는 없다. 번역에 관한 철학적이고 관념적인 접근 방식이 있을 수 있으며, 실무적 차원에서의 다양한 접근 방식이 있을 수 있기 때문이다. 예를 들어 20세기 프랑스 번역학의 토대를 마련한 연구자들 중의 한 명인 조르주 무냉은 경험적 지식보다는 번역의 언어적·과학적 분석을 시행하였다(신정아 2009; 윤성우 2009). 그럼에도 불구하고 최소한 번역 현상 자체에 대한 문제에 있어서는 직간접적으로 경험적 태도를 배제할 수는 없을 것이다. 그러한 연구는 번역 자료들을 직접적으로 분석하는 것이 중요하기 때문이다.
- 7) 1세대 자동번역의 실패는 번역을 형식 문법에 의한 언어의 규칙화된 이동으로 생각했기 때문이다.

찰이 중요하며, 이 사례의 보고인 번역 코퍼스는 번역 연구에서 핵심적인 역할을 할 수가 있는 것이다.

2.2. 코퍼스: 정성과 정량의 교차점

방금 언급한 것처럼 번역학에서 코퍼스를 사용하는 것이 어느 정도 필연적인 결과임을 인식했다면, 이제는 과연 번역 연구를 위해서 이 코퍼스를 어떻게 활용을 해야 하는지에 관한 문제가 제기된다. 흔히 코퍼스 기반 연구라고 했을 때, 가장 먼저 연상이 되는 것은 컴퓨터, 전자문서, 통계 분석과 같은 용어들이다. 조준형(2014)에서 언급하고 있듯이 넓은 의미에서의 코퍼스는 ‘텍스트 자료의 집합’이다. 어떤 형태이든 언어 자료를 담고 있는 텍스트는 모두 코퍼스라고 할 수 있다. 그럼에도 불구하고 코퍼스는 곧 컴퓨터 처리라는 인식이 생겨난 것은 아마도 코퍼스의 개념이 시간이 흐를수록 전문화되고, 또 언어자동처리 분야에서 이를 많이 활용하면서부터라고 할 수 있다. 현대 코퍼스는 ‘언어 표본으로서 전자문서 형태의 텍스트 언어자료’로 정의되기 때문이다. 이는 번역 코퍼스로 마찬가지로 할 수 있다. 그렇지만 모든 번역 연구자가 컴퓨터에 익숙한 것도 아니며, 모든 코퍼스 기반 연구가 정량적 태도로 이루어지는 것도 아니다.

일반적으로 코퍼스 기반 번역 연구의 출발점으로 모나 베이커(Mona Baker)의 1993년 논문이 언급된다. 이 논문에서 베이커는 코퍼스 언어학에서 개발된 기법들을 번역 연구에 어떻게 적용할 수 있고, 코퍼스가 번역 원리의 규범화를 찾아내기 위해 얼마나 유용할 수 있는지를 주장하고 있다. 그 이후, 베이커(1995), 요한슨(Johansson 1994), 라비오사(Laviosa 1997), 케니(Kenny 1998) 등 다양한 연구가 뒤를 잇고 있다. 이와 함께, 기디언 투리(Gideon Toury)의 기술번역학(Descriptive Translation Studies)도 코퍼스 기반 번역 연구를 촉발시키는 원인이 되고 있다.

라비오사(2002: 5-22)에 따르면, 기술번역학과 코퍼스 기반 번역학은 서로 유사점이 있으면서도 차별화되는 부분이 있다. 우선 두 분야 모두 번역 연구에서 코퍼스의 활용을 주장하는데, 이는 번역 원리의 단순화(simplification), 보편화(universalization) 그리고 규범화(norm)를 이끌어 내는 데 코퍼스가 중요한 역할을 할 수 있기 때문이다. 번역 사례를 담고 있는 코퍼스의 관찰에 의한 경험

적 연구를 바탕으로 번역 과정에서 나타날 수 있는 보편적 원리를 발견할 수 있다는 것이다.⁸⁾ 그런데 앞서 언급한 것처럼 기술번역학에서 말하는 경험적 연구는 코퍼스가 포함하고 있는 번역 사례들에 대한 인문학적 해석 및 비판, 즉 정성적 분석이 핵심적인 연구 태도이다. 반면, 코퍼스 기반 번역 연구는 기존의 코퍼스 언어학에서 개발된 기법들을 그대로 수용하여 번역어의 빈도 조사 그리고 번역에 관련된 확률적 계산을 동시에 활용하고 있다. 정량적 분석에 해당하는 이러한 연구 방법은 코퍼스 기반 언어학에서는 빈번하게 사용된 연구 기법이라고 할 수 있다.

번역 연구에서 코퍼스를 기반으로 연구를 한다는 것은 일차적으로 코퍼스에서 번역 대응 관계를 이루는 쌍을 추출해낸다는 것을 가리킨다. 그렇다면 ‘어떤 자료를 어떻게 추출해낼 것인가?’라는 의문이 생긴다. 전통적인 연구 방식이라고 한다면 연구자가 직접 문헌 자료, 즉 번역 텍스트를 관찰하면서 필요한 자료를 추출해낼 것이다. 이러한 연구 태도에서는 복잡한 정량 분석을 기대하기 어렵기 때문에, 정성적 분석 위주가 될 가능성이 높다고 할 수 있다. 반면 현대적 코퍼스 기반 연구는 번역 탐색을 위해 컴퓨터를 이용하거나 전문분석도구를 사용하게 되는데, 이 도구들은 일단 코퍼스를 개별 단어 단위로 분리하고 나서 각 어휘들의 빈도수를 계산하고 이를 토대로 다양한 탐색 기법을 제공하게 된다. 따라서 자연스럽게 정량분석을 동반하게 된다.

그렇지만 코퍼스를 기반으로 하더라도 반드시 전문분석 도구를 사용해야 한다는 것은 아니며, 컴퓨터를 활용하더라도 정량적 분석을 시행하지 않을 수도 있다.⁹⁾ 그리고 분석 대상에 따라서 정량적 분석보다는 정성적 분석이, 혹은 그 반대의 분석 태도가 필요할 수 있다.

번역은 하나의 언어 기호로 표상된 의미를 다른 언어 기호로 재표현하는 것을 의미한다. 여기서 중요한 것은 형태 차원의 이행이 아니라, 언어 기호라는

8) 기술번역학에서 규범은 선형적으로 존재하는 것이 아니며, 오로지 경험적 연구에 의해 후천적으로 규정되는 개념이다. 그리고 규범의 보편적 특징을 위해서 코퍼스는 대규모 코퍼스를 사용해야 하며, 소규모의 개별 코퍼스를 가지고 이러한 규범을 규정하는 것은 어렵다고 주장하고 있다(라비오사 2002: 5-22 참조).

9) 최희경(2015)에 따르면, 번역학에서 코퍼스 기반 연구라고 하더라도 실제 분석은 정량적 분석보다는 정성적 분석 태도를 취하는 연구가 더 많다.

의복으로 감싸진 의미가 다른 언어 체계에서는 어떤 식으로 재포장되는지를 확인하는 것이다. 그래서 번역사적인 측면에서 주요 번역가들은 번역에서 중요한 것은 형태가 아니라 의미라고 주장해 왔다. 현대 번역학에서도 우리가 주목해야 하는 것은 바로 이런 측면이며, 현재 많은 연구자들이 번역의 다양한 현상을 관찰하고 이를 일반화하고자 시도하고 있다.

이러한 관점에서 번역 연구를 위해 코퍼스에 기대하는 것은 풍부한 번역 자료, 더욱이 고품질의 번역 자료이다. 연구자들은 바로 이 고품질의 번역 자료를 코퍼스에서 찾아내고 이를 다양한 연구에 활용하고자 하는 것이다. 연구자의 이러한 목표를 충족시키기 위해서 잘 구축된 코퍼스가 요구되는 것이다. 그렇다면 ‘번역 연구를 위한 코퍼스는 항상 고품질의 번역 텍스트로만 구성이 되어야 하는가?’라는 의문이 제기될 수 있다.

번역 연구에서 코퍼스가 중요한 것은 이 도구가 번역의 실제 현실을 반영하고 있기 때문이다. 코퍼스가 번역의 이상적 상태를 제시하는 것이 아니라, 우리가 실제로 어떻게 번역이 이루어지는지를, 다시 말해서 전문 번역가이든 비전문 번역가이든, 번역가들이 어떤 식으로 번역을 하고 있는지를 코퍼스를 통해서 관찰할 수 있기 때문이다. 그래서 이를 통해서 번역 대응관계, 언어 간의 차이, 번역의 품질평가 등 다양한 정성적 분석을 할 수 있기 때문이다. 이런 점에서 반드시 고품질의 번역 텍스트로만 코퍼스를 구축하는 것은 오히려 코퍼스 연구의 본질을 왜곡할 위험이 있다. 코퍼스의 주요한 이점들 중 하나는 바로 번역 오류를 관찰할 수 있다는 점이다.

코퍼스로 선택된 번역 텍스트는 항상 좋은 번역만을 포함하는 것이 아니다.¹⁰⁾ 코퍼스는 다양한 언어 현실을 반영하며, 여기에는 비문법적 표현도 포함될 수 있다. 번역의 관점에서도 잘못된 번역, 다시 말해서 번역 오류도 포함된다. 그리고 번역 오류는 연구자가 오류가 발생한 이유와 더불어 번역 대안을 제시할 수 있는 원천이 될 수 있다. 그래서 번역 오류는 단순히 배제해야 할 것이 아니라 발전된 번역을 위한 발판이 될 수 있다. 교육적인 차원에서도 이러한 오류들은 번역 훈련을 위해 좋은 예들이 될 수 있다. 정성적 분석의 관점에서 좋

10) 번역 코퍼스 구축에서 너무 많은 번역 오류를 가진 텍스트는 분명 배제되어야 할 것이다. 번역 오류라고 하더라도 학술적인 차원에서 그리고 실용적인 차원에서 가치가 있는 것들이 필요하기 때문이다.

은 번역이든 잘못된 번역이든, 이 모든 예들이 번역의 개념 및 심층적 원리를 설명하기 위한 표본들이 될 수 있다. 코퍼스에서 추구해야 할 것은 바로 이러한 점들이며, 코퍼스는 이러한 연구에 대한 명시적인 예들을 제공할 수 있어야 한다. 그렇기 때문에 코퍼스를 구축한다는 것은 단순히 이런저런 번역 텍스트를 수집한다는 일차원적인 차원의 작업이 되어서는 안 된다. 따라서 코퍼스를 구축한다는 것부터가 정성적 태도를 요구한다고 할 수 있다.

그런데 언어 연구에서도 컴퓨터가 활용되기 시작하면서 정성적 분석-엄밀한 의미에서 인간 연구자의 정성적 분석이라고 할 수 있다-과 함께 다양한 통계적 처리가 동반되는 정량적 분석의 비중도 점점 더 높아지게 된다. 베이커의 논문과 이후 코퍼스 기반 인문학적 번역 연구가 등장하기 시작한 1990년대 초에 언어자동처리 분야에서도 이와 관련된 연구가 시작되었는데, 특히 번역 코퍼스로부터 자동으로 번역어들을 추출하는 것과 관련된 연구들이 주를 이루었다. 이 연구들은 병렬코퍼스의 자동 구축과 관련이 되었으며 서로 다른 언어로 작성된 원문과 번역문 사이에서 문장 수준의 번역 정렬을 위해서 통계적 자료들을 주요 변수로 고려하였다. 브라운 외 2인(Brown *et al.* 1991), 지미나(Zimina 2004), 게일과 처치(Gale & Church 1993) 등이 대표적인 예라고 할 수 있다. 이자벨과 워릭-암스트롱(Isabelle & Warwick-Armstrong 1993)은 베이커 및 투리와 마찬가지로 번역 연구에서 코퍼스 사용을 강조했지만, 핵심은 자동번역을 위한 연구였다. 반면, 베이커(1995)와 라비오사(2002: 23-31)는 번역 연구에서 코퍼스 기반 언어학의 방법을 활용할 것을 제안하고 있지만, 번역 맥락을 관찰하기 위해서 키워드 검색 방식을 예로 든 것이지 번역의 정량화가 연구의 목표는 아니었다.

비록 순수 언어이론 연구자들은 언어라는 인문학 자료를 정량적 분석의 대상으로 두는 것을 반대하고 있지만, 응용언어학, 언어 산업 등 언어학 분야가 확대되면서 정량적 연구를 무시할 수만은 없게 되었다. 이것은 번역학에서도 마찬가지인데, 기계번역을 고려한다면 번역의 대응 가능성이 높은 번역쌍을 찾아내어 이를 자동번역에 반영한다면 보다 높은 품질의 번역 결과물을 획득할 수 있을 것이다. 여기서 번역의 정량적 분석이 중요해진다. 그래서 번역 산업을 고려한다면 이제 정량적 분석 자체를 무조건 배제할 수만은 없을 것이다. 그렇지만 중요한 것은 앞서 언급한 것처럼 번역 연구의 본질을 외면해서는 안 된다

는 것이다. 이창수(2011), 조준형(2013a, 2013b, 2013c)은 번역 연구에서 정량적 분석의 방법을 사용한 연구이지만, 이 연구들의 궁극적인 목적은 정량적 분석의 당위성을 주장하는 것에 있는 것이 아니라 코퍼스에서 번역 현상 관찰(예를 들어, 공기어, 반복구문, 구 단위의 번역대응)을 위해서 정량적 분석이 어떻게 적절한 증거를 제공할 수 있는지를 보여주는 데 있다. 우리가 번역 연구에서 우선적으로 고려해야 할 코퍼스의 개념은 번역 사례를 담은 텍스트 집합이라는 그것의 본질적인 특성이지 정량화 자료로서의 데이터가 아니다.

3. 정성과 정량의 경계

3.1. 정성인가? 정량인가?

코퍼스 기반 연구에서 정량 분석의 주목적은 대규모 코퍼스에서 번역 대응 가능성이 있는 번역쌍을 효율적으로 찾아내기 위해서이다. 그래서 어휘들의 출현 빈도를 고려해서 번역쌍을 찾아내거나, 어휘그룹 내에서 개별 단어들 간의 연관관계를 비교해서 번역 대응 가능성을 추적한다. 이를 위해서 어휘탐색을 시행하고 그 결과물인 어휘맥락색인(concordance)을 바탕으로 맥락 양상을 서로 비교하여 어떻게 번역이 되고 있고, 어떤 오류가 나타나는지를 검증하는 방식이 활용된다. 그리고 t 검정을 비롯한 다양한 통계기법을 사용하여 어휘들의 상관성을 살펴보기도 한다. 더욱이 실무적인 차원에서 번역 연구는 설문조사 기법이 많이 활용되기 때문에 통계나 시각화 도구(그래프와 도표)가 동반될 수밖에 없으며, 이러한 정량적 자료들에 의해 설문조사의 결과가 보다 명시적이고 체계적으로 제시될 수 있다.

그런데 여기서 우리가 주의해야 할 것은 번역 연구를 위해 사용되는 통계 기법들이 어디까지나 확률적인 수치라는 것이다. 확률이라는 것은 하나의 가능성을 의미한다. 따라서 확률적으로 번역 대응 가능성이 높다고 하더라도 항상 그렇게 번역이 되어야만 한다는 것을 의미하는 것은 아니며, 마찬가지로 확률 수치가 낮다고 하더라도 그렇게 되어서는 안 된다는 것을 의미하는 것은 아니다. 통계 수치는 한정된 코퍼스에서의 확률적 결과물일 뿐이다. 그래서 코퍼스

가 달라지면 통계 수치는 달라질 수가 있다. 그래서 저빈도 혹은 확률값이 낮다고 하더라도 이 변수를 무조건 무시하는 것 또한 경계를 해야만 할 것이다. 물론 연구를 할 때는 결과의 일반화를 위해서 고빈도 혹은 확률값이 높은 변수들을 연구 대상으로 하는 것이 일반적이다. 한 어휘 혹은 번역쌍이 고빈도라는 것은 코퍼스 내에서 그 형태가 많이 출현한다는 것을 의미하며 결과에 대한 신뢰성도 그 만큼 높아질 수 있다. 높은 확률값도 같은 의미로 해석할 수 있지만, 이것은 출현 빈도가 높다는 것을 의미하는 고빈도와는 다르게 해석을 해야 할 것이다. 확률값을 계산하기 위해 사용된 변수들은 여러 가지 요소가 포함되는데, 특히 코퍼스의 크기에 따라서 확률값은 달라질 수 있기 때문이다. 저빈도어라고 하더라도 특정 확률값은 고려된 변수에 따라서 높게 나타날 수도 있다. 그래서 확률적 통계 수치에 대한 해석은 정량적 분석과 동시에 정성적 분석도 수반되어야 한다.

앞서 코퍼스는 정성분석과 정량분석의 교차점이라고 말하였다. 실질적으로 정성적 태도로 코퍼스를 연구해야 인문학적인 가치가 있다고 단언하기도 어려울뿐더러, 코퍼스 기반 연구에서는 반드시 정량분석이 포함되어야 한다고 말하기도 어렵다. 왜냐하면 많은 번역 연구에서 정량적 자료가 포함되지 않는 연구도 많기 때문이다. 중요한 것은 어떤 분석 관점을 가져야하는 것이 아니라 자신의 연구에 맞게 두 가지 태도를 적절하게 조화를 이루도록 하는 것이 타당할 것이다. 이제 그 이유를 살펴보기로 하자.

3.2. 코퍼스에 따른 정량적 자료의 변화

김정우(2013)의 연구에서도 알 수 있듯이, 코퍼스 크기가 일정한 크기를 넘어서면 반복되는 번역어 요소들이 일정해지기 때문에 정량적 자료들이 정성적 분석에 미치는 영향이 점차 줄어든다. 따라서 무조건적으로 코퍼스의 크기를 늘리는 것은 통계적인 측면에서 잉여적일 수 있는 결과를 보일 수 있기 때문에 무의미할 수도 있다. 코퍼스 크기에 대한 명확한 규정이 없는 것도 바로 이런 이유 때문이라고 할 수 있다. 그렇다고 하더라도 그것은 균형 코퍼스처럼 코퍼스 구성을 최적으로 했을 때 가능하며, 코퍼스를 구성하는 텍스트의 종류가 특정한 주제에 치중되어 있을 경우는 문제가 있을 수 있다. 해당 주제에 관련된

번역어의 정량적 자료는 일정한 특징을 가지지만, 그렇지 않는 번역어들은 전체적인 번역적 특징을 규정하기가 힘든 수치를 보일 수가 있다. 그래서 균형 번역코퍼스를 구축하는 것은 매우 신중하게 수행되어야 할 작업이다.

그렇지만 개인 연구자들이 자신의 연구를 위해 사용하는 코퍼스는 물리적인 한계 때문에 소규모로 구축되는 경우가 일반적이다. 이 경우에 코퍼스에 포함된 번역어들의 수가 한정되어 있고 특히 이 번역어들이 반복되는 양상이 대규모 코퍼스와는 달리 선택된 텍스트에 따라서 다른 빈도 및 통계적 특징으로 나타날 수 있기 때문에 정량적 평가에서는 매우 신중한 접근이 필요할 것이다.

예를 들어, 영국의 소설가 조앤 K. 롤링(Joan K. Rowling)의 해리포터(Harry Potter) 1-2부를 코퍼스로 하여 번역 연구를 한다고 가정해보자.¹¹⁾ 이 소설에 포함된 어휘들 중 인명을 제외하고 가장 높은 빈도를 보이는 보통명사는 ‘문’을 의미하는 *door*이다. 번역적 판단을 위해서 *door* 단일어보다는 일정한 어휘그룹(cluster)을 조사하는 것이 나올 수 있기 때문에, *door*가 포함된 세 개의 어휘를 포함하는 어휘그룹을 추출해보면 다음과 같다.

〈표 1〉 해리 포터 1-2부에서 *door*를 포함한 반복구문

1부		2부	
빈도수	어휘그룹	빈도수	어휘그룹
16	on the door	6	at the door
14	to the door	6	on the door
10	the door and	6	opened the door
10	the door open	5	the door and
10	through the door	5	the door of
8	pulled the door	4	closed the door
8	the front door	4	open the door

11) 해리포터 소설을 분석하기 위해서 분석도구인 *AntConc*를 사용하였다. 해리포터 1부인 『해리포터와 마법의 돌(Harry Potter and The Sorcerer's Stone)』은 80,624의 어형을 가지고 있으며, 2부인 『해리포터와 비밀의 방(Harry Potter and the Chamber of Secrets)』의 88,143의 어형을 가진 코퍼스이다. 두 소설의 코퍼스 크기는 거의 유사하다고 볼 수 있다.

6	slammed the door	4	pushed the door
6	the compartment door	4	the front door
6	the door, and	4	to the door

1부와 2부에서 *door*가 포함된 어휘그룹의 출현 형태를 비교하면 전체적으로 비슷한 양태를 보이고 있지만, 몇 가지 점에서 차이가 있다. 예를 들어, 전치사를 포함하는 어휘그룹, *on the door*, *to the door*가 서로 다른 빈도 특징을 보이고 있으며, *at the door*는 2부에서는 고빈도로 나타나지만, 1부에서는 도표에 포함되어 있지 않다.¹²⁾ 그리고 *slammed the door*는 1부에서는 빈도수 6을 가지지만, 2부에서는 빈도수 1을 가진 어휘그룹이다. 빈도수를 기준으로 판단하면, *slammed*와 *the door*는 1부에서는 두 어휘의 연계 연관성이 높은 하나의 의미단위로 고려할 수 있지만, 2부에서는 우연적 결합이라고 판단할 수도 있다. 조금 극단적인 예이긴 하지만, 이처럼 같은 주제의 소설에서도 텍스트가 달라짐에 따라서 통계적 결과가 달라질 수 있기 때문에, 완전히 다른 주제의 텍스트를 선택할 때는 정량적 수치에 대해 더욱 신중한 분석이 필요할 것이다.

다른 예를 보면 *open*이라는 동사가 1부에서는 *door* 오른쪽에 있지만, 2부에서는 왼쪽에 위치한다. ‘The’, ‘door’, ‘open’ 세 가지 어휘의 구성이지만 1부와 2부에서 통사적 구조가 다르며, 이는 번역에서도 서로 다르게 실현될 수도 있다. 해당 맥락은 다음과 같다.

the door open
 Griphook held *the door open* for them.
 Harry pulled *the door open* and they ran inside.
 They pushed *the door open*.

open the door
 I'll be waiting to *open the door*.

12) *Door*를 포함한 어휘그룹 전체를 고려하면, *at the door*는 빈도수 4를 가진 어휘그룹이다. <표 1>은 *door*를 포함한 어휘그룹들 중에서 빈도수 그리고 철자순서를 기준으로 상위 10개의 어휘그룹만을 제시하고 있기 때문에 *at the door*가 누락되어 있다.

Even someone invisible would've had to *open the door*.
 Watching Harry fearfully, Neville pushed *open the door*.

실제 맥락을 고려하면 *open*의 위치는 언어적인 차원에서 큰 의미를 가지지 못한다. 따라서 정성적인 사유 없이 단순히 분석 도구에서 추출된 결과물 그리고 빈도수만을 가지고 어떤 번역적 혹은 언어적 분석을 한다면 위에 제시된 예에서처럼 전혀 다른 판단을 할 수 있다.

3.3. 고빈도 자료와 저빈도 자료

코퍼스 기반 번역 연구에서 원문과 번역문 사이의 번역 대응 관계를 평가하기 위해서 고려하는 정량적 변수들 중 가장 기본이 되는 것이 바로 빈도수이다. 코퍼스에서 빈도수를 알아내기 위해서 전문 분석도구를 많이 사용하는데, 이때 분석도구는 코퍼스로 사용된 텍스트를 어형 단위로 해체하면서 어형 목록과 빈도수를 함께 제시한다. 이때 연구자가 빈번하게 범할 수 있는 오류가 바로 빈도의 오류라고 할 수 있다. 사실 빈도수는 번역의 상관성을 표상하는 확률값과 달리, 단순히 코퍼스 내에서 해당 어휘의 출현 횟수를 가리킨다. 다시 말해서 코퍼스에서 A라는 단어가 총 10번 출현하면, A의 빈도수는 10이 된다.

그런데 저빈도어의 경우에 이것은 분석 코퍼스에서 상대적으로 출현빈도가 낮다는 것을 의미하는 것이지 절대적으로 그렇다는 것을 나타내는 것은 아니다. 그리고 번역 텍스트를 보면 일반적으로 많이 사용되는 어휘들은 그 만큼 빈도수가 높게 나타날 것이며, 반대로 잘 쓰이지 않는 어휘는 빈도수가 낮게 나타나는 것은 당연하다. 특히 문학 작품에서 지역색을 나타내는 문화어들과 전문 용어들, 신조어들은 주제에 따라서 빈도수가 달라질 수밖에 없다. 이런 차원에서 저빈도어는 번역의 결정적인 지표가 될 수도 있다. 조금 더 극단적으로 말하면 코퍼스 전체 혹은 세분화된 하위 코퍼스에서 빈도수가 1인 형태(*hapax legomena*)는 해당 코퍼스 혹은 하위 코퍼스의 표지가 될 수 있다. 실제로 언어 자동처리 분야에서 병렬코퍼스를 자동으로 구축하는 방법을 연구하는 연구자들은 이 빈도수가 1인 형태를 번역 문장을 연결하기 위한 결정적인 표지로 고려하는 경우도 있다. 고빈도어들은 코퍼스 내에서 어휘들의 결합형태가 다양하기

때문에 번역 대응을 결정하기 위해서는 많은 정보를 고려해야 하지만, 저빈도어들은 코퍼스 내에서 특정 어휘와 결합할 가능성이 높기 때문에 저빈도어 자체가 번역 표지 역할을 할 수 있다. 그래서 지미나(2002), 조준형 & 이영훈(2012)처럼 저빈도어를 활용한 번역 대응 연구를 하는 경우도 있다. 이 연구들에 따르면 저빈도어들이 무조건 무의미한 것이 아니라 단지 해당 코퍼스에서 출현빈도가 낮을 뿐이지, 번역과 관련해서는 유의미한 자료로 텍스트에 따라서 번역적 판단을 위한 중요한 표지가 될 수 있다.

분명히 정량적 분석은 코퍼스에서 정성적 분석이 간과하기 쉬운 번역 특징들을 찾아낼 수 있는 중요한 분석 태도이며, 특히 코퍼스 기반 연구나 자동번역 분야에서는 최소한의 필요충분조건이라고 할 수 있다. 그렇지만 위에서 언급한 것처럼 통계 자료는 코퍼스로 사용된 텍스트에 따라서 달라질 수 있기 때문에, 이에 대한 해석은 신중해야 할 것이다. 따라서 정량적 분석 결과는 정량적 해석으로 그치는 것이 아니라, 이 과정에서도 정확한 가치판단을 위해서 정성적 해석이 요구된다고 할 수 있다.

3.4. 정량의 정성적 분석

지금까지 우리는 정량적 분석의 몇 가지 문제점을 살펴보았다. 그렇다고 해서 코퍼스 기반 연구에서 정성적 분석만이 타당하다고 주장하기는 어렵다. 작은 크기의 번역코퍼스를 사용한다면 연구자는 코퍼스를 면밀하게 관찰하면서 번역 현상을 연구할 수 있지만, 대규모 코퍼스에서는 여러 가지 측면에서 인간 연구자가 직접 코퍼스를 관찰하는 것은 매우 비효율적이다. 따라서 후자의 경우에는 컴퓨터의 이용하는 것이 효율적이며, 특히 특정 어휘의 전체적인 번역적 특징을 설명하기 위해서는 정량적 자료들이 필요하다. 이때 중요한 것은 정량적 판단에 대한 정성적 분석이 수반되어야 한다는 것이다. 이것은 정량적 자료들과 실제 번역과의 차이를 보일 때 그 원인을 설명하는 것이 중요하기 때문이다. 다음 도표는 영어 *white collar*와 이 표현의 프랑스어 번역을 함께 제시하는 병렬코퍼스이다.¹³⁾

13) <표 2>는 대표적인 번역코퍼스 사이트인 *Linguee*(<http://www.linguee.fr>)를 참조하였다.

〈표 2〉 *White collar*의 영어-프랑스어 병렬코퍼스

영어	프랑스어
As we have already said, we will take other [...] measures to punish white collar criminals.	Comme nous l'avons déjà mentionné, nous prendrons d'autres mesures pour [...] sévir contre les criminels en col blanc .
The salary of white-collar workers is set individually and separately, which means that no pay comparisons can be made.	Les fonctionnaires perçoivent des salaires individuels et différenciés, ce qui signifie qu'il n'y a aucune comparaison des salaires.
Three out of four white collar workers are computer users.	Trois cols blancs sur quatre utilisent un [...] ordinateur.
I am proud to say that yesterday this government tabled legislation that will [...] crack down on white collar crime.	Je suis fier de dire que, hier, le gouvernement a présenté une mesure législative qui sévira plus [...] sévèrement contre les crimes conomiques .
The security of employment and the pension schemes also [...] worsened and the staff got new [...] agreements either for white collar or blue collar workers, [...] depending on what kind of new job they got.	La sécurité de l'emploi et les systèmes de retraite se sont également détériorés, et [...] le personnel a reçu de nouvelles [...] conventions, soit pour cols blancs , soit pour cols bleus, selon [...] le type d'emploi obtenu.
The Government is also helping to make communities safer from violent criminals and predators by introducing [...] important legislation that will protect children from online predators and protecting all [...] Canadians from white-collar crime.	Le gouvernement aide également à protéger les communautés contre les criminels violents et les dangereux prédateurs en prenant [...] d'importantes mesures législatives contre les prédateurs qui s'attaquent aux enfants en [...] ligne et les criminels en col blanc .
My father lost his white-collar job and became a bricklayer.	Mon père perdit son emploi de bureau et devint maçon.

〈표 2〉에서 우리는 영어 *white collar*가 프랑스어로 얼마나 다양하게 번역되어 있는지 확인할 수 있다. 코퍼스 기반 번역 연구는 우선 코퍼스에서 영어

*white collar*가 포함된 맥락과 이에 대응되는 프랑스어 번역 맥락을 추출한 후에 두 언어 사이의 번역 관계를 설명하는 것으로 이루어질 것이다. 정성적 분석의 관점에서는 두 텍스트 각각의 맥락과 해당 어휘의 의미 관계를 동시에 고려해서 영어 *white collar*를 중심으로 한 번역 과정 그리고 최종 결과물을 정리하는 것이 주요 분석 방법이 될 것이다. 인간 연구자라면 위 텍스트를 직접 관찰하면서 이러한 작업을 수행할 수 있다. 만일 코퍼스의 크기가 커진다면 이러한 작업은 많은 시간과 노력을 요구한다. 따라서 컴퓨터를 이용하는 것이 효과적일 수 있다. 컴퓨터를 활용하여 코퍼스를 분석하는 경우는 인간 연구자와는 다른 방식으로 코퍼스에 접근할 수밖에 없다. 번역 텍스트를 분석할 때 인간 연구자는 직관적으로 맥락 전체를 순차적으로 확인하면서 작업이 이루어지지만, 컴퓨터는 이러한 방식을 수행할 수 없기 때문에 다음 몇 가지 다중적인 단계가 필요하다.

1. 원문과 번역문 사이의 번역 맥락 인식 - 인간 혹은 기계에 의한 선작업에 의해 가능¹⁴⁾
2. 대응된 번역 맥락에서 어휘들의 분포를 비교.
3. *White*와 *collar*의 연관성 비교.
4. 영어 원문 맥락과 프랑스어 번역본 맥락에서 가능성이 있는 번역어 추출.

위 작업에서 특히 3-4 작업이 진행되는 과정에서 컴퓨터는 통계적 정보를 기반으로 번역 연관성을 추정하게 된다. 우리는 이 과정에서 인간 연구자와 기계 사이의 차이를 분명하게 확인할 수 있다. 먼저, 3에서 *white*와 *collar*의 연관성, 다시 말해서 두 어휘가 직접적인 의미 단위를 형성하는지 그렇지 않은지를 그리고 어떤 의미인지를 인간 연구자는 바로 확인이 가능하지만, 컴퓨터는 이러한 판단을 직접적으로 할 수가 없다. 일차적으로 컴퓨터는 두 어휘가 동일한 맥락(일반적으로 문장)에서 함께 출현한다는 것만을 알 수 있다. 위 도표의 예에서처럼 *white*와 *collar*가 바로 인접해서 출현하더라도 컴퓨터는 이 그룹이 하

14) 일반적으로 많은 시간과 비용을 요구하는 이 작업은 병렬코퍼스 구축을 위한 필수적인 선행 작업이다.

나의 의미 단위를 이루는지 아니면 단순한 결합인지를 판단할 수가 없다. 예를 들어 *White collar and pink sweater with a diamond necklace* 맥락에서는 <표 2>의 *white collar*와는 다른 의미를 가진다.

4단계에서 인간 연구자는 텍스트 관찰과 동시에 번역어라는 것을 인지할 수 있지만, 컴퓨터는 그렇지 못하다. 이때 빈도수와 같은 정량적 자료가 판단 기준이 될 수 있다. 그렇지만 이것은 모두 경우에 적용될 수 있는 것은 아니다. 예를 들어 위 도표에서 *white collar*의 빈도는 7이지만, 프랑스어에서 직역이라고 할 수 있는 *col blanc*은 4의 빈도를 보인다. 적은 수치에서는 어느 정도 수용 가능하지만 만일 코퍼스의 크기가 커지면서 두 어휘 사이에서 상대적 출현 비율의 격차가 커진다고 한다면 두 표현 사이의 번역 가능성을 의심할 수밖에 없을 것이다.

위의 <표 2>를 보면 영어 *white collar*의 프랑스어 번역 대응어로서 *col blanc*, *fonctionnaires*, *économique*, *de bureau* 세 가지 형태가 나타나고 있다. *Col blanc*과 달리 나머지 두 표현은 빈도수 1로 나타나는데, 빈도수만을 번역 평가의 변수로만 고려한다면, 나머지 세 가지 표현은 두 언어 사이의 번역 대응이 무시될 가능성이 매우 높다. 더욱이 *fonctionnaires*와 *de bureau*와 달리 *économique*는 일반적인 상황에서 *white collar*의 번역어가 될 가능성이 거의 없다고 할 수 있다.¹⁵⁾ 그렇지만 확률적으로 낮다고 하더라도 *white collar*가 *économique*로 번역되는 사례가 있다면 그리고 이것이 오역이 아니라면, 이것은 번역적으로 중요한 하나의 사례가 될 수 있다. 이 경우 연구자는 정성적 태도를 가지고 그렇게 번역된 이유를 찾아내어야 할 것이다. 우리의 예를 보면 프랑스어 *économiques*로 번역된 맥락을 보면 이 형용사는 *crime*이라는 프랑스어 명사와 결합되어 있고, 영어 원문에서도 *white-collar crime* 형태로 되어 있다. 다시 말해서 ‘범죄’와 결합되어 ‘경제 범죄’로 번역이 되어 있으며, 어떤 의미에서는 해당 맥락에서는 이러한 번역이 더 나올 수 있을 것이다.¹⁶⁾ 이러한 과정을 통해서 연구자는 영어 *white collar*의 프랑스어 번역을 다양하게 제시할 수 있고,

15) 프랑스어 *économique*는 영어 *economic*에 해당한다.

16) 분석의 엄밀성을 위해서 또 다른 맥락을 *Linguee*에서 검색하면 *crime*이 있는 맥락에서 실제로는 프랑스어 *col blanc*의 비중이 높았고, 소수에 불과하지만 *économiques*의 예도 여러 개 확인할 수 있었다.

어떤 맥락에서 그러한 번역이 가능한지를 체계적으로 설명할 수 있을 것이다.

정량적 분석 없이도 이러한 예들을 찾을 수 있겠지만, 정성적 태도와 정량적 태도의 순환적 분석은 특히 코퍼스 기반 연구에서 인간 연구자의 눈에서 간과될 수 있는 부분들을 보완할 수 있는 효과적인 태도가 될 수 있다. 이러한 관점에서 코퍼스 기반 번역 연구에서 정량적 형식화의 오류를 해결하고, 정성적 분석의 부족함을 보완하기 위해서 두 가지 관점의 교차적인 활용이 필요하다고 할 수 있다.

4. 결론

번역 사례를 기반으로 한 번역 연구에서 코퍼스는 매우 중요한 개념이 되고 있다. 번역코퍼스를 구축한다는 것은 현실적으로 아주 힘든 작업이지만, 그럼에도 불구하고 현대 번역학에서 풍부한 번역의 사례를 제공한다는 차원에서 매우 유용한 도구가 된다. 우리는 본 논문에서 과연 코퍼스가 정성 분석의 대상인가 아니면 정량 분석의 대상인가에 대한 질문을 던지고 이에 대한 답을 찾고자 하였다. 사실 코퍼스 기반 연구라고 하더라도 번역, 특히 인간의 언어가 관련된 연구에서는 정성적 태도가 본질이 되어야 할 것이다. 그렇다면 질문은 정량적 태도가 번역 연구에서 반드시 필요한가라는 질문으로 수정된다.

넓은 의미에서 번역코퍼스를 텍스트 번역 자료라고 한다면, 이것을 가지고 번역을 연구하는 것은 기존의 연구방식과 큰 차이가 없다고 할 수 있다. 실제 번역 연구에서 코퍼스에 대한 정량적 분석 없이 정성적 분석으로 이루어지는 연구가 매우 많다. 그렇지만 대규모 코퍼스에서 번역의 양상에 대한 전체적인 특징을 찾아내기 위해서는 컴퓨터라는 도구를 활용할 수밖에 없고, 특히 자동 번역 분야에서는 코퍼스에서 번역 대응어들의 연결 관계를 찾을 수 있는 매개 변수들이 필요하다. 이 경우에 확률값과 같은 정량적 자료들은 필수적이라고 할 수 있다.

앞서 살펴본 것처럼 서로 보완적인 관계 속에서 정성적 자료와 정량적 자료를 고려하는 것이 코퍼스 기반 번역 연구가 효과적으로 이루어질 수 있도록 하기 위한 필요충분조건이라고 하였다. 중요한 것은 자신이 연구하는 주제

에는 어떤 분석적 태도가 필요한지를 연구자가 정확하게 인식하는 것이다. 그리고 정량적 자료들을 있는 그대로 제시하는 것으로 연구를 마무리하기 보다는 이 자료에 대한 면밀한 정성적 해석이 이어지는 것이 필요하다. 이러한 순환적 방식에 의해 인문학자들, 특히 언어학자들에 의해 비판받는 언어 자료의 계량화가 정당한가라는 비판에서 자유로울 수 있으며, 정량적 자료들도 비로소 의미 있는 가치를 획득할 수 있게 될 것이다.

참고문헌

- 강지혜 (2013) 「경험적 연구로서의 번역사례연구: 방법론적 고찰」, 『번역학연구』 14(2): 7-38.
- 김정우 (2013) 「코퍼스기반 번역학 연구에서 정량적 인자가 정성적 분석 결과에 미치는 영향」, 『번역학연구』 14(1): 83-98.
- 김화영 (1999) 「번역이란 무엇인가?」, 『미메시스』, 창간호, 18-35.
- 신정아 (2009) 「조르주 무냉(G. Mounin)의 번역론에 관한 소고 - 『레 벨 앵피 델』을 중심으로」, 『번역학연구』 10(1): 31-101.
- 윤성우 (2009) 「조르주 무냉(George Mounin) 번역론의 몇 가지 철학적 전제들에 관한 소고」, 『번역학연구』 10(1): 103-20.
- 이영훈 (2013) 「한국번역학사 기술을 위한 전제와 시론」, 『번역학연구』 14(2): 187-222.
- 이창수 (2011) 「엔그램(n-gram) 분석을 통한 번역한국어와 비번역한국어간의 어휘묶음 유형 차이 사례 연구」, 『통번역학연구』 15(1): 317-40.
- 정호정 (2003) 「코퍼스 중심의 번역학 연구」, 『번역학연구』 4(2): 71-88.
- 조재룡 (2009) 「번역사를 바라보는 한 관점: 앙리 메쇼닉(H. Meschonnic)의 경우」, 『번역학연구』 10(1): 141-70.
- 조준형 (2012) 「병렬코퍼스에서 맥락 탐색의 의미와 한계」, 『번역학연구』 13(5): 223-46.
- 조준형 (2013a) 「불한 번역 코퍼스에서 반복구문 분석에 의한 번역 대응 연구」, 『번역학연구』 14(4): 163-98.

- 조준형 (2013b) 「불한 병렬코퍼스에서 다중공기어에 의한 번역 대응 연구: 고빈도어를 중심으로」, 『통번역학연구』, 17(3): pp. 359-86.
- 조준형 (2013c) 「비교코퍼스를 활용한 한불-불한 번역 연구」, 『통역과 번역』, 15(1): 257-88.
- 조준형 (2014) 「코퍼스 기반 번역 연구의 인식론적 고찰」, 『번역학연구』, 17(3): 279-301.
- 조준형, 이영훈 (2012) 「『모데라토 칸타빌레(*Moderato Cantabile*)』의 번역코퍼스에 나타난 저빈도 명사 연구」, 『통역과 번역』, 14(1): 237-71.
- 최희경 (2015) 「코퍼스를 활용한 번역연구」, 『2015년 한국번역학회 봄 학술대회 발표논문집』, 57-62.
- Baker, Mona (1993) 'Corpus Linguistics and Translation Studies: Implication and Application', in *Text and Technology : In Honour of John Sinclair*, 233-50.
- Baker, Mona (1995) 'Corpora in Translation Studies : An Overview and Some Suggestions for Future Research', *Target*, 7(2), 223-43.
- Baker, Mona (ed). (1998) *Routledge Encyclopedia of Translation Studie*. London & New York: Routledge.
- Ballard, Michel (1992) *De Cicéron à Benjamin. Traducteurs, traductions, réflexions*. Presses Universitaires de Lille.
- Brown, Peter. F., Jennifer. C. Lai & Robert L. Mercer (1991) 'Aligning Sentences in Parallel Corpora', *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*: 169-76.
- Gale, William A. & Kenneth W. Church (1993) 'A program for aligning sentences in bilingual corpora', *Computational Linguistics*, 19(1): 75-102.
- Holmes, James S. (1972) The Name and Nature of Translation Studies, *Translated!: Papers on Literay Translation and Translation Studies*, Amsterdam: Rodopi: 67-80.
- Isabelle, Pierre & Susan Warwick-Armstrong (1993) 'Les corpus bilingues: une nouvelle ressource pour le traducteur', Pierrette Bouillon et André Clas (Dirs.), *La Traductique: études et recherches de traduction par*

ordinateur, Les Presses de l'Université de Montréal: 288-306.

Laviosa, Sara (2002) *Corpus-based Translation Studies: Theory, Findings, Applications*, Amsterdam & New York: Rodopi.

Toury, Gideon (1995) *Descriptive Translation Studies and beyond*. Amsterdam/Philadelphia: John Benjamins.

Zimina, Maria (2002) Repérages lexicométriques des équivalences à basse fréquence dans les corpus bilingues. *Lexicométrie, no spécial*, 1-27.

[Abstract]

**What is the corpus in translation studies?:
qualitative analysis or quantitative analysis**

Cho, Joon-Hyung
(Korea University)

What is the corpus ? A *corpus* is the textual sample of linguistic data produced by a language community. In the same way, we can define the *translation corpus* as a collection of translations data. The Descriptive Translation Studies and the Corpus-based Translation Studies insist on a corpus-driven research into translations for the universalization of translation principles, because a translation corpus contains the samples of translations produced by professional or non professional translators.

But the corpus is considered as a computational tool in the field of Machine Translation, and so the corpus-based research usually accept quantitative analysis. Strictly speaking, the translation is human language activity and the place where two different language-culture meet. Therefore it is a phenomenon of human sciences. And then the main goal of a corpus-based translation studies is to explain the fundamental principle of translation depending on a parallel corpus. Consequently the translation corpus must be considered as a textual data that allows to systematize explicitly translation phenomena and its theory (qualitative analysis), not merely a formal data (quantitative analysis).

▶ Key Words: corpus, qualitative analysis, quantitative analysis, translation corpus, translation studies, translation

조준형

고려대학교 불어불문학과 강사

chojh4net@naver.com

관심분야: 코퍼스 번역학, 번역평가

논문투고일: 2015년 5월 3일

심사완료일: 2015년 5월 31일

게재확정일: 2015년 6월 8일