

## 코퍼스 기반 번역 연구의 인식론적 고찰

조 준 형  
(고려대)

### 1. 서론

현대 언어 관련 연구 분야<sup>1)</sup>에서 많이 활용되고 있는 ‘코퍼스’는 이질적인 학문들의 교차점이 되면서 인문학적인 측면(언어 텍스트)과 통계·공학적인 측면(전자문서, 표본, 통계, 컴퓨터 처리)이 혼재된 개념으로 정의된다. 시간이 지남에 따라서 개념은 확대되고 학자들마다 정의하는 방식은 차이가 나지만, 그 출발점은 ‘인간 언어활동의 결과물인 언어 텍스트의 집합’이며, 이것이 ‘코퍼스’의 가장 핵심적인 개념이라고 할 수 있다.<sup>2)</sup>

번역 연구는 현재 언어뿐만 아니라 영상, 이미지, 기호 등 다양한 기호학적

- 
- 1) 언어 관련 연구는 순수 언어학뿐만 아니라 응용 언어학 그리고 언어 산업이 모두 포함된다.
  - 2) ‘언어 텍스트 집합’이라는 기본 개념에 통계적 방법과 전산언어학 기법들이 적용되면 서, 코퍼스의 정의에 ‘표본’, ‘통계처리’, ‘전자문서’, ‘데이터베이스’ 등의 개념들이 포함되었다고 할 수 있다.

요소들을 포함해서 번역의 개념 자체에 대한 철학적·인식론적인 성찰, 그리고 문화, 정치, 경제와 같은 사회적 원리에 이르기까지 매우 다양한 주제들을 대상으로 하고 있다. 그러나 번역은 근본적으로 서로 다른 언어-문화권 사이의 의사소통을 목적으로 하는 것이기 때문에 하나의 언어 체계에서 다른 언어 체계로 이행되는 과정에서 발생하는 원리를 규명하는 것을 출발점으로 한다. 이러한 관점에서 연구자들이 번역 현상의 관찰 자료로서 코퍼스에 주목하는 것은 매우 자연스러운 현상이라고 할 수 있다.

그런데 번역학의 연구 주제가 다양해짐에 따라서 코퍼스 역시도 구어 혹은 문어의 순수 언어 텍스트뿐만 아니라 이미지, 녹음된 음성, 도상 등 다양한 요소들이 코퍼스의 대상이 되었으며, 더욱이 번역 코퍼스도 2개의 언어를 대비시키는 것뿐만 아니라 그 이상의 언어로 구성된 다국어 코퍼스도 등장하게 되었다. 이와 함께 분석 방식도 통계학 및 전산언어학의 기법들을 응용하면서 보다 다양한 분석 결과를 제시할 수 있게 되었다. 문제는 코퍼스와 관련된 이러한 발전 양상이 기술의 발달과 함께 너무나 전문화되다보니 오히려 인문학연구자로서의 번역학 연구자들이 코퍼스에 쉽게 접근할 수 없을 현상이 나타나게 되었다. 코퍼스 연구는 통계처리와 컴퓨터 처리를 동반하고, 이에 대한 전문지식이 없다면 코퍼스와 관련된 연구는 힘들다는 인식이 은연중에 배어있는 것 같다.

현대 번역학을 비롯한 언어 연구에서 ‘코퍼스’의 활용이 강조되고 있고, 모나 베이커(Mona Baker)나 미셸 발라르(Michel Ballard) 등과 같은 학자들이 코퍼스의 개념을 중시하는 것은 결코 이 도구의 전문화된 분석 방법들 때문이 아니라 바로 이 도구가 가지는 핵심적인 개념 ‘인간 번역활동의 실제 결과물’ 때문일 것이다. 바로 번역의 실제 결과물을 분석함으로써 번역 연구자는 언어 간의 번역 이행 과정에서 실제로 어떤 원리가 작동하고 있는지를 이론적이 아닌 경험적·실험적으로 확인할 수 있는 단초를 코퍼스가 제공할 수 있기 때문이다.

그럼에도 불구하고 국내 번역학계에서 코퍼스는 여전히 한편으로는 필요하지만 다른 한편으로는 불편한 도구로 받아들여지고 있다. 이것은 코퍼스 분석이 매우 전문화된 태도를 요구하며, 이에 따라 코퍼스 처리의 외형적인 측면만이 부각되기 때문인 것 같다. 그렇지만 코퍼스라는 개념 자체의 본질을 이해한다면 번역 연구에서 이 도구의 사용에 대한 주저함은 불필요한 행위일 뿐이다.

이러한 관점에서 본 논문은 코퍼스라는 하나의 개념을 인식론적인 관점에서 코퍼스의 본질은 무엇이고, 왜 이를 번역 연구에서 활용해야 하는지를 살펴보고자 한다. ‘무엇이 코퍼스이고, 왜 코퍼스인가?’라는 코퍼스의 근원에 관한 문답을 통해서 번역이 단순히 실무차원의 응용 연구에만 국한되는 것이 아니라 오랜 시간 이루어져 온 인간 언어활동이며, 코퍼스가 이를 연구하는 데 있어서 중요한 바탕이 될 수 있음을 탐구해 나갈 것이다.

## 2. 코퍼스의 본질

통역이 서로 다른 언어-문명권 사이의 의사소통을 위해 탄생한 인간의 언어 활동이라고 한다면, 번역은 의사소통과 더불어 또 다른 목적성을 가지는 보다 복합적인 언어활동이라고 말할 수 있다. 역사적으로 번역은 관점의 차이는 있었지만 동시대의 발달된 사상과 문화를 수용하거나 확산시키는 일종의 통로였다. 특히 “유럽은 번역으로부터 그리고 번역에서 태어났다”는 메쇼닉(Meschonic 1999: 38)의 말처럼 유럽에서 국가와 언어의 형성에서 번역은 중요한 역할을 하였으며, 언어적·문화적·정치적·경제적·민족적 전략이 함의되어 있었다.<sup>3)</sup> 번역사를 이야기할 때 항상 가장 먼저 등장하는 성경 번역은 히브리어-그리스어(70인역, BC2세기)-라틴어(성 히에로니무스(St. Jérôme)의 불가타(Vulgate) 성경, 4세기)-프랑스어를 비롯한 유럽어로의 번역의 이동이 일어나는데, 이것은 곧 해당 시기에서 유럽의 중추가 되는 문화 그리고 국가-민족어로 발돋움하고자 하는 언어들의 성장과 일부 관련이 있다고 할 수 있다. 그리고 아랍의 수학, 의학, 천체와 관련 문서들의 번역도 역시 아랍 문명의 위상을 확인할 수 있다.

3) 유럽인의 시각에서 문화와 번역 사이의 관계를 기술한 코르도니에(Cordonnier 1995)의 『번역과 문화(Traduction et culture)』에 따르면 고대부터 20세기 이전까지 번역은 발달된 문명의 전파 경로가 되었지만, 유럽의 민족-국가 및 민족-언어의 신장과 교양을 위한 하나의 수단으로서 타자를 배제하고 자아중심의 민족주의적인 성향을 띄게 되었다. 이것은 비유럽 국가에 대한 유럽 국가들의 침략의 역사 그리고 언어적 폭력으로 이어지게 된다.

이 모든 흐름에는 강성한 문화에 의한 강력한 언어의 힘이 존재한다. 그리고 후진 문명은 발달된 문명의 지식을 번역을 통해 수용하면서 다른 한편으로 자국어의 어휘력과 표현력을 높이려고 노력하였다. 번역은 두 언어의 충돌을 전제하기 때문에 어느 한 언어의 변질을 가져올 수 있다. 번역사에서 자주 등장하는 직역과 의역의 문제는 바로 이러한 점과 관련이 있다고 할 수 있다. 심지어 번역을 통한 이국성을 배제시키고 모국어를 지위를 향상시키고자 했던 의도가 엿보이는 16세기 프랑스어의 번역불가능성(intraduisibilité)도 궁극적인 언어-문화적인 목적과 분리되지 않는다.<sup>4)</sup> 사실 이러한 논쟁은 현재도 지속적으로 나타나고 있으며, ‘직역과 의역’, ‘원천주의자(sourcier)와 목표주의자(cibliste)’ 논쟁은 여전히 해결되지 않고 있는 대표적인 논쟁이다.

번역학은 바로 이 복잡한 번역 현상을 밝히기 위해 탄생한 학문이라고 할 수 있다. 매우 복잡한 문제들이 번역에 관련이 되지만 번역학은 크게 두 가지 범주로 구분될 수 있는데, 한편으로 번역 활동의 기저에 대한 개념을 설명하기 위한 순수 이론적 연구와 다른 한편으로 실제 사례 분석을 통한 번역의 구체적인 원리를 살펴보는 연구라고 할 수 있다. 비록 두 가지 연구 범주로 구분은 하였지만 두 범주는 서로 분리된 것이 밀접한 연관이 있으며 번역학의 발전을 위해서는 두 방향의 연구가 균형 있게 진행되어야 한다.<sup>5)</sup>

코퍼스 기반 번역학은 주로 전자보다는 후자의 연구에 적합한 것이라고 할 수 있다. 사실 코퍼스의 개념에 대해서는 많은 학자들이 정의를 내리고 있지만

4) 16세기 뒤벨레(Du bellay)의해 시작되었다고 알려진 번역불가능성은 사실 프랑스 문학, 특히 시(詩) 분야로 국한된다. 시 문학의 특수성 때문에 언어적인 차원의 번역은 이루어질 수 없다는 것이다. 따라서 번역불가능성은 16세기 모든 종류의 장르에 적용된 것은 아니라고 할 수 있다. 이에 대해서는 조재룡(2009: 157-169), 베이커 편역(1998: 410-411) 참조.

5) 국내 번역학에서는 상대적으로 후자의 연구가 더 활발하고 전자의 연구는 상대적으로 위축되어 있는 것 같다. 이것은 번역에 대한 순수 이론적·사상적·철학적 논의들이 주로 유럽에서 발달이 되었고 유럽어 학제보다는 영미권의 학제 소속의 번역학자들이 더 많은 국내 여건상 연구자들이 유럽어로 작성된 텍스트에 접근하기가 쉽지 않기 때문이다. 이러한 측면에서 윤성우(2009), 윤성우 & 이향(2011), 이향(2011), 전성기(2012; 2013)의 연구는 번역에 대한 순수 이론적 성찰을 이해하는 데 참조할 수 있는 연구라고 할 수 있다.

“언어자료를 모아놓은 뭉치로 우리말로는 말모듬 또는 말뭉치라고도 일컫는다”(김정렬, 이동주, 전희철 2012: 2)는 표현이 코퍼스의 핵심적인 개념을 가장 적절하게 설명하고 있다. 코퍼스는 한국어 표현인 말모듬이다. 다시 말해서 언어 연구를 위해서 언어 텍스트를 모아 놓은 집합체이다. 이를 통해서 우리는 언어의 실상에 접근할 수 있다. 그런데 코퍼스 언어학의 발생 자체가 컴퓨터의 사용과 밀접한 관련을 가지기 때문에 이 정의에, 표본, 전자문서, 컴퓨터 처리 등 매우 전문적이고 기계적인 개념들이 포함되었다. 안동환(2010: 17)은 코퍼스 언어학에서의 모든 문제와 해결은 다음과 같은 질문과 연관되어 있다고 말한다.

- 텍스트를 기계 검색이 가능한 형태로 저장시킴으로써 얻는 이점을 어떻게 가장 잘 활용할 수 있을까?
- 어떤 언어 이론이 코퍼스기반 연구를 구축하는데 가장 도움이 될까?
- 어떤 언어 현상을 찾아야 할까?
- 어떤 응용에 이 연구에서 파생하는 통찰력과 개선된 언어기술을 활용할 수 있을까?

이 네 가지 질문에서 번역의 관점에서 우리가 주목해야 할 것은 바로 세 번째와 네 번째 질문이다. 첫 번째 질문은 코퍼스의 본질에 대한 것보다는 활용적인 측면과 관련되며 두 번째 질문은 사실 번역학의 관점에서는 쉽게 답하기 힘든 질문이라고 할 수 있다. 코퍼스 언어학은 언어학 이론들이 이미 많은 연구가 이루어져있기 적절한 방법론을 찾아내는 것이 상대적으로 쉬울 수 있지만, 현대 번역학에서 이론과 응용에서 어느 쪽이 더 우위에 있는지 말하기 어렵다. 문학과 철학과 관련된 번역을 제외하면 일반적으로 실무적인 성격이 강하기 때문이다.

코퍼스 기반 번역 연구에 가장 적절한 질문은 바로 세 번째 질문이 될 수 있다. 실제로 번역을 연구할 때 가장 중요한 것은 번역 텍스트를 관찰하는 것이다. 번역 텍스트에 나타나는 다양한 번역 용례들을 관찰하면서 연구자는 서로 다른 언어가 만날 때 어떤 현상이 일어나며 어떤 원리가 작동하는지 밝혀낼 수 있다. 앞서 우리는 번역이 단순히 언어적인 문제가 아니라 철학적·교육적·문화적·사회적 문제 등 다양한 요소들과 관련된다고 말하였다. 그런데 이러한 주제들이 번역의 테두리에서 논의될 수 있는 것은 번역 텍스트 자체에서 이 모

든 개념들이 내포되어 있기 때문이며, 번역 텍스트를 배제하고는 이러한 문제를 논의하는 것은 어렵다고 할 수 있다. 다시 말해서 텍스트의 관찰을 통해 방금 언급된 개념들에 대한 논의로 확대될 수 있다.

발라르(2007: 17-18)는 번역 연구에 있어서 가장 핵심적인 것은 바로 언어, 보다 엄밀하게 말해서 언어와 텍스트라고 말하고 있다. 여기서 텍스트를 언급한 것처럼 번역 연구가 언어의 일차적인 의미가 아니라 텍스트라는 거대한 맥락 속에서의 언어를 연구해야 한다는 것을 말한다. 그리고 “코퍼스를 기반으로 한 번역 연구는 원문과 번역문, 두 텍스트의 대조에서 시작한다”는 (발라르 2007: 24)의 언급은 코퍼스를 기계적인 의미의 형식적인 개념이 아닌 텍스트로서의 코퍼스를 지칭하는 것이다. 우리는 바로 이러한 개념으로서의 코퍼스를 바라보아야 할 것이다.

단일어 연구와 번역 연구는 상당한 차이가 있다. 단일어 연구에서도 언어적인 요소 이외에 화행, 문화, 기호학적 양태 등 다양한 개념들이 논의가 되지만, 번역은 텍스트를 통해서 이질적인 두 개의 문화가 충돌하는 것이다. 단순히 언어의 이행이 아니다. 문화와 사상과 철학이 동시에 포함된 텍스트라는 거대한 집단의 이행이다. 인문학적 관점에서 코퍼스도 바로 이러한 맥락에서 이해를 해야 할 것이다.

실상 지금까지의 코퍼스 기반 연구는 이러한 개념을 충족시키기에는 여러 면에서 부족했던 것도 사실이다. 분석 방법이 지극히 형식적이고 정량적 분석에 치중하다보니 정작 중요한 번역에 대한 정성적 분석이 부족했기 때문이다. 그리고 코퍼스 언어학의 영향으로 컴퓨터라는 기계적인 수단 때문에 코퍼스를 기계처리의 자원으로만 고려되었다. 다른 한편으로는 코퍼스 분석은 코퍼스 내의 번역 용례들을 살펴보는 연구가 많았기 때문에 사례 연구에 적합한 것으로 고려되기도 하였고 현재도 그러한 인식이 강하게 나타나고 있다.

그런데 코퍼스에 대한 이러한 인식은 오히려 인문학 연구자들이 이에 대한 심도 있는 성찰이 부족한데서 원인을 찾을 수 있다. 코퍼스의 사용에 적극적으로 기보다는 기계적인 처리의 어려움으로 오히려 코퍼스를 멀리하려고 하지 않았는지 고민을 해보아야 한다.

코퍼스를 구축하고 이를 활용한다는 것은 여러 가지 이점을 가진다. 먼저 국가 차원에서 언어 자원을 확보하는 것이다. 언어는 시대에 따라서 의미와 사

용 조건들이 조금씩 변화한다. 사진은 이러한 변화를 즉각적으로 반영하기 힘들고 더욱이 많은 용례들을 담아내기가 어렵다. 두 번째로 구축된 코퍼스를 통해서 연구자들이 실용적인 연구가 가능하다. 코퍼스의 가장 중요한 이점이라고 할 수 있다. 세 번째로 표준적으로 구축된 코퍼스는 다른 영역으로 확장이 될 수 있다. 이는 곧 다양한 언어 관련 분야에서 언어 자원의 재활용이 가능하다는 것이다.

번역학의 관점에서 이 이러한 코퍼스의 이점은 그대로 적용할 수 있다. 대규모 번역 코퍼스가 구축되어 있다면 단일어 코퍼스보다 훨씬 다양한 정보를 담은 언어 자원이 된다. 원문이 가지는 언어-문화적 특징뿐만 아니라 번역문의 언어-문화적 특징을 동시에 포함하고 있기 때문이다. 그리고 언어 표현이 아무리 다양하더라도 궁극적으로는 반복적인 패턴을 가지고 있기 때문에 번역 차원에서 반복되는 표현을 재활용한다는 장점을 가진다. 자동번역에서 ‘번역 메모리’의 개념이 바로 여기서 비롯된다. 이를 통해 번역가가 반복되는 표현을 번역 지원도구를 통해서 번역 자원의 능력을 올릴 수 있으며, 시간도 줄일 수 있다.<sup>6)</sup>

이러한 관점에서 코퍼스는 연구 차원에서뿐만 아니라 실무적인 측면에서도 중요한 언어 자원이 된다고 할 수 있다. 이를 위해서는 잘 구축된 번역 코퍼스가 필요하다. 번역 코퍼스 구축과 관련해서 언어자동처리 분야의 연구자들이 보이는 관심은 어떻게 가장 빠르고 정확하게 병렬코퍼스를 구축할 수 있는가의 문제이다. 따라서 번역 코퍼스로 선택될 텍스트는 어떠한 것이든 상관없다. 그러나 앞서 언급한 코퍼스의 이점을 극대화하기 위해서는 코퍼스로 사용될 번역 텍스트의 선택이 중요해진다. 오류가 적은 번역, 다양한 장르의 텍스트 선택, 표준이 될 수 있는 훌륭한 번역문의 선택과 평가 이것이 바로 코퍼스 기반 번역 연구에서 번역 현상 자체의 연구 이외에 번역학자들이 해야 할 몫이다.

6) 문학 번역은 다분히 문학적 텍스트가 가지는 특징을 고려해야하기 때문에 자동번역에 의한 문학 텍스트 번역은 아직은 불가능하지만, 최소한 실무 텍스트들의 번역에서는 자동번역 혹은 번역지원도구에 의한 번역이 현재 많이 상용화되어 있다고 할 수 있다.

### 3. 인문학에서 코퍼스 기반 번역 연구의 태도

‘코퍼스 (기반) (번역)연구’, ‘코퍼스에 의한 000 번역 분석’, ‘000 코퍼스를 중심으로’. 이러한 표현은 코퍼스와 연관된 연구들에서 흔히 볼 수 있다. 그런데 여기서 ‘코퍼스 (기반) 연구’라는 것은 도대체 어떤 연구 태도를 의미하는가? 중의적인 이 표현에 대해서 우리는 두 가지 연구 태도를 고려할 수 있다. 하나는 ‘코퍼스를 활용한 연구’이며, 다른 하나는 ‘코퍼스를 대상으로 한 연구’이다. 코퍼스와 연관된다는 점에서는 서로 공통점을 가지고 있지만, 전자는 코퍼스가 다른 연구를 위한 수단으로 이용된다는 것을 의미하고, 후자는 코퍼스 자체가 연구의 목표가 된다는 점에서 차이가 있다고 할 수 있다.

이러한 이중적인 태도는 안동환(2008: 17)이 지적하듯이 “코퍼스기반 번역학에 영향과 자극을 준 2가지 주요 원천은 코퍼스언어학과 기술번역학(Descriptive Translation Studies: DTS)”이기 때문일 수도 있다. 그리고 안동환(2008: 19)은 “모든 코퍼스언어학적 분석은 연구대상인 코퍼스 자체의 생성과 그리고 그것을 관찰하고 분석하고 처리할 소프트웨어 도구의 개발”에 달려있다고 언급하고 있다. 이를 코퍼스 기반 번역학에 그대로 적용을 한다면, ‘모든 코퍼스 기반 번역 분석은 번역 코퍼스 자체의 생성과 그것을 관찰하고 분석할 수 있는 소프트웨어 도구 개발에 있다’라고 말할 수 있다. 그러나 이러한 설명은 코퍼스 기반 연구에 대한 잘못된 이해를 불러올 수 있는 매우 한정된 시각이라고 할 수 있다. 다시 말해서 위의 설명은 ‘코퍼스를 활용한 연구’보다는 ‘코퍼스를 대상으로 한 연구’에 가깝다고 할 수 있기 때문이다.

번역 관련 연구에서 프랑스어에는 *traductologie*와 *traductive*라는 두 가지 용어가 함께 사용된다. 영어의 *translation studies*에 해당하는 *traductologie*는 우리가 알고 있는 번역학을 의미한다. 이와 달리 *traductive*는 영어의 *machine translation*에 대응되는 표현으로 한국어로 ‘기계번역학’<sup>7)</sup>이라고 표현할 수 있을 것이다. 기데르(Guidère 2010: 133-135)에 따르면, 기계번역학은 번역 활동에 전산언어학의 개념을 도입한 것으로 다국어 맥락에서 인간 언어 번역을 목적으

7) *Traductive*를 ‘기술번역학’이라고 옮길 수도 있지만, ‘기술’은 중의적인 표현으로 DTS의 기술번역학을 가리킬 수도 있기 때문에, 본 논문에서는 의미의 혼란을 피하기 위하여 기계번역학이라고 할 것이다.

로 하는 언어자원과 프로그램들을 분석하고 개발하는 것을 목표로 한다. 그리고 그 목적은 “번역작업들을 부분적이든 전체적이든 자동화”(기테르 2010: 134) 하는 것이다. 이러한 관점에서 코퍼스의 구축과 프로그램 개발은 번역학의 영역이 아니라 기계번역학의 영역이라고 할 수 있다. 사실 이 영역에서 인문학 관점에서의 번역연구자들이 할 수 있는 역할은 매우 제한된다. 아마도 번역연구자들이 기여를 할 수 있다면, 코퍼스 구축에 있어서 구성의 문제 혹은 코퍼스로 사용되는 번역텍스트들에 대한 품질 평가 그리고 개발된 프로그램에 대한 평가 같은 부분일 것이다.

그렇기 때문에 기계번역학적 연구는 대부분은 전산언어학자, 컴퓨터 공학자 혹은 통계학자들에 의해 수행된다. 인문학자들도 이 분야에서 연구를 진행할 수도 있지만 컴퓨터에 대한 전문지식을 필요로 하기 때문에 상대적으로 참여가 제한적일 수밖에 없다. 국내에서도 번역학회의 『번역학연구』를 비롯하여 번역 관련 학술지에 실린 논문들을 살펴보다라도 프로그램 개발과 코퍼스 구축 기법에 대한 논문들은 전무하다고 할 수 있다. 이것은 국내 번역학 학술지들이 추구하는 목적이 순수 ‘기계번역학’의 주제들이 아니기 때문이다. 이창수(2011)와 조준형(2013a) 등 전문화된 코퍼스 활용에 관한 연구들이 있지만, 이들 연구 역시도 통계 기법 자체가 아니라 해당 기법 통해 어떻게 번역 현상을 설명할 수 있는지 그리고 그 기법을 번역 연구에서 적절하게 적용시킬 수 있는지가 주요 목적이라고 할 수 있다.

코퍼스 기반 번역 연구를 논의하면서 흔히 언급되는 연구자가 모나 베이커와 기디언 투리(Gideon Toury)이다. 베이커(1995, 1998)는 코퍼스 분석의 한 예로서 키워드 검색에 의한 어휘맥락색인(concordance) 기법을 설명하면서 다양한 유형의 번역 코퍼스(병렬 코퍼스, 비교 코퍼스)가 번역 현상의 원리를 설명하는데 원천이 될 수 있으며, 코퍼스가 번역 연구에서의 유용한 자원이 될 수 있음을 설명하고 있다. 투리(1995)도 기술번역학에 대한 논의를 진행하면서 기술번역학의 원래의 목적인 번역의 규범(norme)과 보편성 그리고 용인성을 설명하는데 코퍼스를 활용한다. 엄밀한 의미에서 기술번역학은 방금 언급한 개념을 설명하기 위한 수단으로서 코퍼스를 기술하는 것이지 이것 자체가 목적이 아니다. “Toury의 모형을 형성하는 원리들과 코퍼스언어학을 형성하는 원리들 사이에는 상이점들도 있다”(안동환 2008: 31)는 점에 주목해야 한다. 코퍼스언어학이 언

어학적이든 전산학적이든 코퍼스 자체의 분석과 탐색에 기반을 둔다면, 기술번역학은 “번역행위를 지배하는 규범들을 발견하기 위한 언어 외적 자료들에 관한 연구가 허용된다”(안동환 2008: 32)는 차이가 있다. 사실 이들 연구들에서 기계번역학의 주제는 거의 언급되지 않는다. 다시 말해서 베이커와 투리 그리고 기타 번역학자들에게 있어서 코퍼스는 도구이지 그 자체가 목적은 아니라고 할 수 있다.

기계번역학에서도 번역 현상의 보편성 그리고 규범은 중요한 문제이다. 그렇지만 이 개념에 대한 투리의 시각과 기계번역학자들의 시각은 연구의 목적에 있어서 많이 차이가 있다고 할 수 있다. 넓은 의미에서 투리를 비롯해서 번역연구자들이 보편성과 규범을 찾아내고자 하는 것은 그 의미 그대로 번역 현상의 자체의 문제와 관련된다. 그리고 코퍼스는 번역가에 의한 실제 번역의 원천이기 때문에 가치가 있는 것이다. 이와 달리, 기계번역학에서의 보편성과 규범은 이를 통해서 번역을 자동화시킬 수 있는 프로그램을 개발하는 것이 선결 목표이다. 즉 번역에서 무수히 반복되어 나타나는 유형들을 분류하고 이를 기반으로 성능이 우수한 자동번역기 혹은 번역지원도구를 개발하는 것이 이 분야의 핵심 주제라고 할 수 있다. 더욱이 자동번역의 결과물이 현실적이고 실용적이며 수용 가능하도록 하기 위해서 코퍼스를 기본 데이터베이스로 사용을 하는 것이 최근의 경향이다.

그렇다고 해서 순수 번역학과 기계번역학을 완전히 구분 지어야한다고 단언하기는 어렵다. 코퍼스를 기반으로 한다는 것 자체가 적용 범위에 상관없이 이미 두 영역을 함께 포용하고 있다는 것을 의미하기 때문이다. 베이커와 투리 이후에 나온 무수한 관련 연구들을 보면 기계번역학의 성과물인 코퍼스 분석도구들, 특히 키워드 검색과 어휘맥락색인과 같은 매우 단순한 기법들을 번역 코퍼스 탐색의 도구로 사용하고 있음을 확인할 수 있다.<sup>8)</sup> 그렇기 때문에 두 영역을 완전히 구분하는 것은 무의미하다고 할 수 있다. 이러한 관점에서 코퍼스 기반 번역학은 ‘코퍼스언어학’과 ‘기술번역학’의 영역의 교차점이 아니라 코퍼스를 교집합으로 하는 ‘기술번역학’과 ‘기계번역학’의 교차점이라고 하는 것이 더

8) 조준형(2013b, 2013c), 베넷(Bennett 2010), 찰스(Charles 2012), 이자벨과 워릭-암스트롱(Isabelle & Warwick-Armstrong 1993)에서 알 수 있듯이 언어와 번역 교육 그리고 번역 연구 등 다양한 주제들에서 키워드 검색과 어휘맥락색인이 사용되고 있다.

타당할 것이다.)<sup>9)</sup>

여기서 현대 학문들의 경향인 학제적 경계를 넘어선 융합 연구의 일면을 찾아볼 수 있다. 그리고 기계번역학과의 만남을 통해서 이미 코퍼스 기반 번역 연구는 기존의 순수 번역 연구, 다시 말해서 순수 이론 연구와는 차별화될 수밖에 없다. 순수 이론과 응용 연구(코퍼스) 모두 번역 연구라는 공통점을 가지지만 후자는 코퍼스의 효과적인 처리와 탐색을 위해서 기계번역학의 성과물들을 사용하는 것이 일반적일뿐만 아니라, 구체적인 분석을 위해서 수학적·전산적·통계적 기법들을 빈번하게 응용한다. 그런데 바로 이러한 특징이 많은 인문 번역학 연구자들 그리고 통번역대학(원)에서 관련 학위논문을 준비하고 학생들이 코퍼스에 쉽게 접근할 수 없도록 만드는 일종의 장벽이 되기도 한다. 현재 국내 외국어 교육을 비롯해서 번역학 관련 학제에서 코퍼스와 관련된 교육은 거의 이루어지지 않고 있다.<sup>10)</sup> 그래서 이에 대해 익숙한 소수의 연구자들을 배제하고, 대다수의 연구자들은 사실상 번역 사례 연구를 진행하면서도 코퍼스 사용에는 어려움을 호소하고 있다.

번역연구자들이 기계번역학의 모든 개념들과 기술적인 방법들을 습득해야 하는 것은 아니다. 그렇지만 코퍼스 기반 번역 연구가 번역학 내에서 하나의 주된 흐름이라면 연구자들은 이러한 환경에 적응을 해야만 할 것이다. 발라르, 베이커, 투리, 이들 학자들이 기계번역학에 대한 기술적인 지식을 완벽하게 가지고 있었다고 말할 수는 없다. 그들은 코퍼스의 가치를 이해하고 이를 탐색하기 위한 최소한의 기술적인 방법론을 활용할 뿐이다. 더욱이 코퍼스 연구가 번역 연구의 모든 문제들을 해결할 수 있는 것은 아니다. 언어 그 자체만이 아니라 비언어적인 요소 그리고 전체적인 맥락의 이해가 필요한 문학 번역, 철학 번역 등과 같은 영역에서 코퍼스 연구는 매우 제한적일 수밖에 없다.

중요한 것은 연구자가 자신의 주제에 적절한 방식을 선택하는 것이며, 코퍼스는 많은 연구 태도들 중의 하나일 뿐이라는 것을 분명하게 인식하는 것이다. 번역연구에서 반드시 코퍼스를 사용해야한다는 강박 관념에 가질 필요도 없으

9) 안동환(2008)에서의 코퍼스언어학은 기계번역학에 속한다고 할 수 있다.

10) 국내 통번역대학(원)에서는 코퍼스 자체에 대한 교육뿐만 아니라 구축, 활용, 전문분석 도구의 사용법 등 코퍼스와 관련된 전반적인 부분에 특화된 교육을 찾아보기가 어렵기 때문에 신진연구자나 학생들 스스로가 이에 대한 지식을 습득해야 한다.

며, 반대로 코퍼스에 대한 무조건적인 거부감을 가질 필요도 없다. 연구의 주체가 번역 용례들을 관찰하거나, 텍스트의 전체적인 맥락을 검토하거나 아니면 대규모의 자료를 분석해야만 하는 경우에는 코퍼스의 활용을 충분히 고려할 수 있을 것이다. 연구의 효율성을 위해서도 그렇게 할 필요가 있다. 언어 자체가 아닌 문화 번역 현상을 연구하더라도 문화어들의 번역 양상을 분석해야만 한다면 코퍼스를 활용한 방식이 도움이 될 수 있다. 원문과 번역문이 속한 언어-문화권에서 어떤 표현들이 대응되고 차이가 있는지를 전체적으로 관찰하고 분류할 수 있기 때문이다. 신문 사설 혹은 기사를 기반으로 한 정치적 사상의 경향을 파악하거나 사회적 분위기를 비교하는 경우에도 마찬가지다. 해당 주제와 관련된 신문 사설들을 단어, 구, 표현 등 다양한 수준에서 통계적 분포와 함께 관찰할 수 있기 때문이다.

이 경우에 연구자는 어떤 형태로 코퍼스를 구축하고, 어떤 통계기법과 전문 분석도구를 사용해야만 한다. 그리고 이러한 선택은 연구의 신뢰성과도 직접적으로 관련된다고 할 수 있다. 적절한 도구의 사용은 코퍼스 연구의 연구 능력을 높이고 결과의 체계적이고 입체적인 설명을 제시할 수 있다. 반면에 부적절한 도구와 분석 기법의 사용은 오히려 겉은 화려하지만 속 내용은 빈 무용한 논의로 비난받을 수 있다. 그럴 경우는 차라리 고전적인 번역 연구 방식으로 회귀하는 것이 나을 것이다. 시간은 오래 걸리지 모르지만 의미적으로 완벽한 텍스트를 직접적으로 관찰하면서 분석 상의 오류는 최소화할 수 있을 것이다. 형식적인 규칙과 프로그램의 효율성이 중요한 기계번역학 연구가 아닌 이상, 코퍼스 기반 연구에서도 가장 무게를 두어야 할 것은 번역 현상 그 자체이기 때문이다.

#### 4. 코퍼스 연구의 신뢰성을 위하여

인문학적인 범주에서 코퍼스 연구를 할 때 가장 주의해야 할 것은 분석 도구, 분석 방법보다는 코퍼스 자체라고 하였다. 분석 도구와 분석 방법은 어떤 의미에서는 연구의 효율성과 분석의 적절성과 관련이 될 수 있지만, 코퍼스 연구 자체의 신뢰성을 확보하기 위해서 잘 구성된 코퍼스가 필요하기 때문이다.

자연과학에서 실험을 할 때 실험 재료 자체가 잘못되어 있다면 그 결과는

신뢰를 받지 못할 것이다. 코퍼스 연구는 용례 관찰을 통한 경험적 연구의 특징을 가지기 때문에, 자연과학과 마찬가지로 관찰의 원천 자료인 코퍼스에 따라서 완전히 다른 결과에 도달할 수 있다. 더욱이 코퍼스 연구는 연구의 주제에 따라서 ‘정량적’ 분석이 동반되기 때문에 ‘정성적’ 분석의 관점에서는 신뢰받지 못하는 일반화의 오류에 빠질 수 있다. 조준형(2012)에서 살펴보았듯이 코퍼스 분석은 일반적으로 단어 수준으로 완전히 해체된 텍스트 코퍼스를 통계 기법으로 재해석하는 것으로 출발한다. 그래서 의미가 배제된 형식적인 차원의 분석으로 고려될 수 있다. 코퍼스 분석의 이러한 일면이 인문학적인 관점에서는 신뢰를 주지 못할 수 있다.<sup>11)</sup>

번역 현상의 일반화를 위해서는 다양한 장르에 속하는 텍스트의 복합적인 관찰이 요구된다. 이러한 관점에서 코퍼스의 크기와 구성이 문제가 될 수 있다. 사실 코퍼스의 신뢰성을 위해서 가장 많이 언급되는 것이 ‘대표성’(représentativité)과 ‘균형’(équilibre)이다. 안동환(2010: 91-96)과 맥어너리, 씨아오 & 토노(McEnery, Xiao & Tono 2006: 15-19)가 언급하듯이, 코퍼스 분석 결과의 일반화를 위해서 다양한 환경에서의 언어 사용을 포함할 필요가 있다. 일반 언어학 연구를 위해서 혹은 단일어 연구에서도 보편성을 확보하기 위해서는 가능한 여러 종류의 텍스트를 고려할 필요가 있다. 그렇다고 모든 텍스트를 코퍼스를 구축하는 것은 현실적으로 어렵기 때문에 대표성을 확보할 수 있는 표본들을 잘 선택하는 것이 중요하다. 그럼에도 불구하고 구축된 코퍼스가 대표성을 가지고 있다고는 말하기 어렵다. 코퍼스는 현실에 존재하는 텍스트의 일부일 뿐이기 때문이다. 더욱이 언어 사용 환경은 텍스트 장르뿐만 아니라 사용자의 조건(연령, 직업, 성별)에 따라서도 변할 수 있기 때문에 이 모든 것을 반영하는 것은 불가능하다고 할 수 있다. 이러한 조건을 번역 코퍼스에 적용하면 더 큰 문제가 나타날 수 있다. 왜냐하면 번역코퍼스는 원문과 번역문이라는 이중적 구조를 고려해야하기 때문이다. 설령 원문의 균형성을 충족하더라도 만일 이 원문에

11) 코퍼스 내 단어 간의 유의미적인 대응 관계들을 측정할 수 있는 기법들(공기어, 반복구문 등)이 있지만 이 역시도 특정 코퍼스에 관련된 수치화된 확률상의 관계만을 보여준다. 따라서 코퍼스를 기반으로 작업을 수행하는 연구자들은 그 결과가 이론 번역학자들의 설득을 이끌어낼 수 있도록 그리고 분석이 형식적인 차원으로만 국한되지 않도록 유념해야 할 것이다.

해당하는 번역문이 존재하지 않는다면 번역문의 균형성은 확보할 수 없으며, 원문적으로 그러한 텍스트는 번역코퍼스에 포함될 수가 없다. 이러한 이유로 번역코퍼스에서 대표성과 균형의 문제는 직접적으로 논의되지 않는 것 같다.

코퍼스의 대표성과 균형의 문제와 함께, 코퍼스의 신뢰성과 관련된 또 하나의 요소는 바로 코퍼스의 크기이다. “아무리 큰 코퍼스라도 주요 언어의 모든 사용자에게 의해 단 하루에 만들어지거나 수신되는 모든 말과 글의 아주 작은 표본보다도 결코 더 많을 수가 없다”(안동환 2010: 97)는 점은 분명하지만 코퍼스의 신뢰성을 위해서는 일정한 크기의 코퍼스는 필요할 것이다. 그럼에도 불구하고 어느 정도 규모 코퍼스가 적절한지에 대해서도 마찬가지로 많은 논란이 있으며, 그 어떤 연구자도 이에 대한 명확한 답변을 주지는 못하고 있다. 엄밀하게 말해서 너무 작은 크기의 코퍼스는 용례의 다양성에 있어서 불충분할 수 있으며, 크기가 작은 만큼 코퍼스에 포함된 텍스트가 특정한 장르에 속할 가능성이 높다고 할 수 있다. 너무 큰 크기의 코퍼스는 자료의 잉여성으로 오히려 작업 효율성을 떨어뜨릴 수 있다. 이에 대해 싱클레어(Sinclair 1991: 9)는 10억 어절의 코퍼스조차도 그리 만족할만한 수준의 크기는 아니라고 설명하고 있지만, 대규모 연구소 수준의 프로젝트가 아닌 이상, 개인 연구자들이 구축하기에는 10억 어절의 코퍼스는 매우 큰 코퍼스이다. 번역코퍼스는 단일어 코퍼스에 비해 구축하는 방식이 매우 상이하고 복잡하기 때문에 개인적인 차원에서 대규모 코퍼스를 구축하기가 용이하지 않다.<sup>12)</sup> 이러한 점에서 코퍼스 크기에 따라서 정성적 결과가 어떤 변화가 있는지를 예증하고 김정우(2013)의 연구를 참조할 수 있다. 이 연구는 일정 분량의 텍스트를 순차적으로 추가하여 코퍼스의 크기가 증가에 따른 정성적 자료들의 변화를 비교하면서 어느 정도의 코퍼스 크기가 분석의 신뢰성을 확보할 수 있는 기준이 될 수 있는지를 검증하고 있다. 물론 이 연구의 결과가 코퍼스의 크기를 결정하는 결정적인 기준이 된다고 단언할 수는 없지만, 개인 연구차원에서는 충분히 참조할만하다고 할 수 있다.

앞서 언급한 것처럼 현실적으로 번역코퍼스는 단일어 코퍼스에 비해 여러 가지 구축의 어려움 때문에 개인 연구 차원에서 대규모 코퍼스를 구축하는 것

12) 번역코퍼스 구성과 구축에 관해서는 김명관(2007: 5-8), 베로니스(Véronis 2000) 참조.

은 거의 불가능하며 균형과 대표성의 확보도 그리 쉬운 문제가 아니다. 연구소 수준의 작업이 아니라면 대규모 번역 코퍼스의 구축은 개인 연구자들이 개별적으로 구축한 소규모의 코퍼스들을 병합하여 하나의 거대한 코퍼스로 구축할 수도 있겠지만, 저작권 문제 그리고 코퍼스를 구축하는 데 개인 연구자가 소비한 노력과 시간을 고려한다면 현실적으로 실현 불가능한 계획이라고 할 수 있다.

실상 순수 언어학과 달리 코퍼스 기반 번역학에서는 자동번역처럼 데이터의 계속된 축적이 자동적으로 이루어지는 경우를 제외하면 대부분의 연구가 소규모 그리고 특정 장르 혹은 작품에서의 번역 문제를 연구하는 경우가 대부분이라고 할 수 있다. 특정 장르의 번역 현상을 관찰한다면 해당 주제를 기술하고 있는 텍스트만을 선택하면 되기 때문에 크기의 문제에 있어서는 상대적으로 자유롭다고 할 수 있다.

아마도 인문학적인 관점에서는 형식적인 문제일 수도 있는 코퍼스의 크기 보다는 코퍼스 자체의 질, 다시 말하면 번역품질에 대한 문제가 더 중요하게 다가올 수 있다. 코퍼스 연구라고 하더라도 번역 연구가 목적이기 때문에 코퍼스로 포함된 번역문 자체가 너무 많은 번역 오류를 포함하고 있거나 혹은 유용한 자료를 추출할 수 없다면 코퍼스의 선택이 잘못되었다고 할 수 있다.<sup>13)</sup> 따라서 연구자가 코퍼스로 사용될 번역 텍스트를 선택할 때, 개략적이라도 사전 검증을 통해서 번역의 오류가 가능한 적은 텍스트, 연구하고자 하는 주제를 검증할 수 있는 용례들이 많이 포함되어 있는 텍스트를 선택해야만 한다. 그렇지 않다면 연구의 신뢰성을 위해서 다른 텍스트로 코퍼스를 대체해야만 할 것이다.

## 5. 코퍼스 기반 번역 연구의 지향점

현대 번역학은 연구 방법과 이론들의 비약적인 발전을 통해 매우 다양한 주제들을 논의의 대상으로 하고 있다. 인문학적인 개념들로만 전개되던 번역 논의들이 이제는 응용 번역학이라는 명칭 하에, 실무적인 차원에서 번역가가

13) 기계번역학의 관점에서는 분석 프로그램의 효율성이 중요하기 때문에 번역 코퍼스의 번역품질은 부차적인 문제일 수 있다. 그래서 만일 원문과 번역문 사이에 오류와 누락이 있다면 이 부분은 텍스트의 맥락과 상관없이 삭제되는 경우도 있다.

활동할 수 있는 모든 영역들이 탐문의 대상이 되고 있다. 기테르(2010)는 번역학 입문서이지만 번역사, 번역 이론, 응용 번역학(영상 문제, 경제, 산업), 번역 교육, 기계번역학, 자동번역 등 매우 많은 주제를 기술하고 있다. 하나의 예에 지나지 않을 수 있지만 실제 번역 연구에서도 이러한 주제들이 활발하게 논의되고 있는 것을 고려한다면, 이제 번역학은 인간의 번역활동이 실행되는 모든 분야들이 번역 연구의 대상이 될 수 있음을 암시한다.

인간 번역활동의 폭이 넓어지면서 번역 텍스트의 장르 및 소재도 마찬가지로 확장되고 있다. 그리고 이를 바탕으로 구축된 코퍼스가 점차 다양해지는 것은 당연한 현상이라고 할 수 있다. 코퍼스에 포함된 텍스트가 다양해지면 이것은 코퍼스에서 하나의 번역어가 가질 수 있는 양태를 여러 환경에서 관찰할 수 있다는 것을 의미한다. 분석할 수 있는 표현들이 다변화되면 번역어의 확장된 사용을 확인할 수 있으며, 분석 결과는 예를 들어 자동번역에 직접적으로 적용하면서 번역 오류를 줄일 수 있는 확실한 자료가 될 수 있다. 그렇지 않더라도 최소한 번역 교육에서 이를 응용하게 되면 통번역을 배우는 학생들뿐만 아니라 실무 통번역가도 이를 적극적으로 활용할 수 있을 것이다. 번역가는 자신의 번역 경험을 늘려가면서 새로운 환경의 번역 상황을 맞이할 수 있다. 이러한 경우에 기존의 연구 결과물들은 번역가에게 매우 실용적인 번역사전이 될 수 있다.

이미 코퍼스언어학뿐만 아니라 번역 코퍼스 연구에서도 순수 언어 혹은 번역 이외의 응용 연구가 활발하게 진행되고 있다. 그리고 그 결과물들은 실용적인 가치가 있는 재활용 자료로 추적되고 있다. 사실상 실무적인 성격이 짙은 국내 번역학 환경에서 순수 철학적·이론적 연구가 아닌 이상은 인문학자로서의 번역 연구가들도 코퍼스가 가지는 실용적인 성격에 주목하여 연구가 그 자체로서의 종결이 아니라 다른 연구자들 혹은 다른 연구 분야에서 그리고 실제 번역가-수요자가 그 결과를 재활용할 수 있는 방법을 모색하도록 노력하여야 할 것이다. 개인 연구자의 입장에서도 하나의 코퍼스 기반 연구를 수행하고 나서 해당 코퍼스를 일회성으로 생각하여 버릴 것이 아니라 전혀 다른 주제를 선택하지 않는 이상 새로운 코퍼스를 기존의 코퍼스에 추가하면서 연구를 심화시킬 수 있다.

아마도 코퍼스 연구는 기계번역학의 영향으로 지금보다 훨씬 더 전문화되고 형식적인 것이 될 수도 있다. 인문학자로서 번역 연구가는 그러한 흐름을 완

전혀 무시해서도 안 되겠지만, 궁극적으로 코퍼스 연구도 결국 번역 연구라는 본래의 목적을 항상 염두에 두어야 할 것이다.

## 6. 결론을 대신하며

이집트 원정을 떠난 나폴레옹(Napoléon) 군대의 장군이었던 피에르 부샤르(Pierre Bouchard)는 1799년 나일강 어귀에서 하나의 비석을 발견한다. 오늘날 로제타석(Rosetta Stone)으로 알려진 이 비석은 이집트 상형문자, 이집트 민중문자, 고대 그리스어로 작성된 일종의 병렬코퍼스였다. 프랑스 언어학자 장-프랑수아 샹폴리옹(Jean-François Champollion)과 영국 언어학자 토마스 영(Thomas Young) 사이에서 프랑스와 영국의 국가 간 자존심 싸움으로까지 이어진 이집트 상형문자 해독에 결정적인 단서를 제공한다. 어떤 의미에서는 극단적일 수도 있는 로제타석은 병렬코퍼스를 가지고 연구자가 인문학적으로 어떤 연구를 할 수 있는지를 보여주는 좋은 본보기가 될 수 있다. 그 당시까지 전혀 이해를 할 수 없었던 이집트 상형문자가 로제타석으로 해독될 수 있었던 것은 비석에 새겨진 텍스트들이 번역 텍스트였기 때문이라고 할 수 있다. 그렇지 않았다면 어휘들의 연결 관계를 확인할 수 없기 때문에 이 문자의 해독에는 조금 더 많은 시간이 필요했을지도 모른다.

지금도 처리하기가 쉽지 않은 일종의 3개어 병렬코퍼스는 코퍼스의 개념이 전혀 없었던 시점에서 이루어진 코퍼스 기반 연구, 더욱이 번역 텍스트이기 때문에 코퍼스 기반 번역 연구라고도 할 수 있을 것이다. 두 언어학자는 로제타석에 새겨진 이집트 민중문자 텍스트와 고대 그리스어 텍스트를 바탕으로 고대 상형문자를 하나하나 대비시키면서 해독 작업을 진행하였다. 다시 말해서 번역 대응이 될 수 있는 어휘들을 찾아내어 이를 하나씩 비교하면서 의미를 파악한 것이라고 할 수 있다. 이러한 작업은 현재 코퍼스 기반 연구에서 번역 대응 관계를 살펴볼 때 연구자가 가지는 연구 방식과 다르지 않다.

만일 지금의 환경이라면 기계번역학 연구자들이 이 텍스트들을 전자문서로 변환하여 컴퓨터를 이용해서 암호 해독기처럼 빠른 속도로 언어 요소들을 대응 시키면서 언어의 비밀을 풀어냈을지도 모른다. 코퍼스라는 개념이 생기기 이전

에도 문헌을 분석하는 것은 이미 언어 및 번역 연구에서 보편적인 방식이었다.

1960년대 컴퓨터의 탄생 시기와 맞물리면서 이를 활용한 언어 연구에 대한 개념을 정의할 필요가 있었으며, 이러한 상황에서 컴퓨터 분석에 적용되는 전자문서 형태의 자료가 코퍼스라는 명칭을 부여받았다고 할 수 있다. 이러한 환경에서 탄생한 코퍼스 언어학은 기존 연구와의 차별화를 위해서 인문학자가 보기에는 매우 형식적이고 기계적인 특성을 가지게 되었고, 이후 코퍼스 언어학이 활성화되면서 코퍼스는 곧 컴퓨터학이라는 오해가 생겨난 것 같다. 특히 번역코퍼스는 베이커가 번역 연구에서 코퍼스 활용의 유용성을 언급하기 이전에 이미 기계번역학의 대상이었다. 단언어 코퍼스 연구자들은 다국어 코퍼스, 정확하게 말해서 다국어 병렬코퍼스에 관심을 가지게 되었고 이를 구축하는 방법을 기계번역학의 관점에서 연구하게 되었다. 이것은 자동번역 혹은 번역지원도구 그리고 번역어 자동추출과 같은 기계번역 개발을 위한 것이었다.

이러한 분위기가 팽배해지면서 번역 연구자들은 베이커나 투리의 주장대로 이 코퍼스의 유용성은 인정하지만 쉽게 다가서지는 못하는 모순된 상황이 나타나게 되었다. 베이커, 투리 그리고 발라르가 코퍼스를 언급한 것은 분명 기계번역학적인 관점에서가 아니다. 이들은 번역 연구에서 코퍼스가 가지는 가치를 인식하고 있었기 때문에 그러한 주장을 한 것이다.

물론 기계번역학의 성과를 완전히 무시하라는 것은 아니다. 기계번역학에서 제안된 도구들 그리고 방법론은 분명히 코퍼스를 체계적으로 그리고 효과적으로 처리하기 위해서는 필요하다. 그러나 그것은 어디까지나 보조적인 수단이며 번역 연구의 목표는 아니다. 우리는 코퍼스의 핵심적인 개념은 ‘언어 자료의 집합체’, ‘실제 언어 사용의 표본’이라고 말했다. 인문학자로서의 번역 연구자는 바로 이점에 초점을 맞추어야 한다. 모두가 기계번역학자가 될 필요는 없다. 번역의 모든 것은 텍스트 그 자체에 있다. 번역학자는 샹폴리옹 그리고 토마스 영이 되어야 한다. 번역학자는 너무 전문적이고 기계적인 성향만을 가질 수 있는 코퍼스를 순수한 번역 연구의 자료로서 균형을 잡을 수 있도록, 그래서 신진 연구자들과 통번역대학원의 학생들이 코퍼스를 쉽게 활용할 수 있도록 노력해야 할 것이다.

## 참고문헌

- 김명관 (2007) 「불-한 병렬코퍼스 구축과 활용」, 『한국프랑스학논집』 58: 1-18.
- 김정우 (2013) 「코퍼스기반 번역학 연구에서 정량적 인자가 정성적 분석 결과에 미치는 영향」, 『번역학연구』 14(1): 31-98.
- 김정렬, 이동주, 전희철 (2012) 『코퍼스 언어학과 영어교육기반 번역학: 이론, 연구결과, 응용』, 서울: 한국문화사.
- 안동환, 옴김 (2008) 『코퍼스기반 번역학: 이론, 연구결과, 응용』, 서울: 동인.
- (Sara Laviosa (2002) *Corpus-based Translation Studies: Theory, Findings, Applications*, Amsterdam & New York: Rodopi)
- \_\_\_\_\_ (2010) 『코퍼스언어학개론』, 서울: 한국문화사.(Graeme Kennedy (1998) *An Introduction to Corpus Linguistics*, London: Longman)
- 윤성우 & 이향 (2011) 「번역인식론적 접근」, 『번역학연구』 12(4): 99-118.
- 이향 (2011) 「번역학과 메타담론」, 『번역학연구』 12(1): 191-210.
- 윤성우 (2009) 「조르주 무냉(George Mounin) 번역론의 몇 가지 철학적 전제들에 관한 소고」, 『번역학연구』 10(1): 103-20.
- 이창수 (2011) 「엔그램(n-gram) 분석을 통한 번역한국어와 비번역한국어간의 어휘묶음 유형 차이 사례 연구」, 『통번역학연구』 15(1): 317-40.
- 조재룡 (2009) 「번역사를 바라보는 한 관점: 앙리 메쇼닉(H. Meschonnic)의 경우」, 『번역학연구』 10(1): 141-70.
- 전성기 (2012) 「인문학 번역론의 한 메타적 성찰」, 『번역학연구』 13(1): 205-40.
- \_\_\_\_\_ (2013) 「번역인문학과 화쟁적 학문하기」, 『번역학연구』 14(1): 275-300.
- 조준형 (2012) 「병렬코퍼스에서 맥락 탐색의 의미와 한계」, 『번역학연구』 13(5): 223-46.
- \_\_\_\_\_ (2013a) 「불한 번역 코퍼스에서 반복구문 분석에 의한 번역 대응 연구」, 『번역학연구』 14(4): 163-98.
- \_\_\_\_\_ (2013b) 「불한 병렬코퍼스에서 다중공기어에 의한 번역 대응 연구: 고빈도어를 중심으로」, 『통번역학연구』 17(3): 359-86.
- \_\_\_\_\_ (2013c) 「비교코퍼스를 활용한 한불-불한 번역 연구」, 『통역과 번역』

15(1): 257-88.

Baker, Mona (1995) Corpora in Translation Studies : An Overview and Some Suggestions for Future Research. *Target*, 7(2), 223-43.

\_\_\_\_\_ (ed). (1998) *Routledge Encyclopedia of Translation Studies*. London & New York: Routledge.

Ballard, Michel (1992) *De Cicéron à Benjamin. Traducteurs, traductions, réflexions*. Presses Universitaires de Lille.

\_\_\_\_\_ (2006) La traductologie, science d'observation, *Qu'est-ce que la traductologie?* Études réunies par M. Ballard, Artois Presses Université: 179-94.

\_\_\_\_\_ (2007) Pour un rééquilibrage épistémologique en traductologie, *Quo vadis Translatologie?*, Frank & Timme: 17-34.

Bennett, Gena R. (2010) *Using CORPORA in the Language Learning Classroom - corpus linguistics for teachers*. The University of Michigan Press.

Cordonnier, Jean-Louis (1995) *Traduction et culture*. Paris: Didier.

Charles, Maggie (2012) Using hands-on concordancing to teach rhetorical functions: evaluation and implications for EAP writing classes, in Ana Frankenberg-Garcia, Lynne Flowerdew & Guy Aston (eds), *New Trends in Corpora and Language Learning*, 26-43.

Guidère, Mathieu (2010) *Introduction à la traductologie. Penser la traduction : hier, aujourd'hui, demain*. Bruxelles: De Boeck.

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.

Meschonnic, Henri (1999) *Poétique du traduire*. Paris: Verdier.

Sinclair, John (1991) *Corpus, Concordance, Collocation*, Oxford: Oxford University Press.

McEnery, Tony, Richard Xiao & Yukio Tono (2006) *Corpus-Based Language*

*Studies*. Routledge.

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

Véronis, Jean (2000). Aligement de corpus multilingues, in Jean-Marie  
PIERREL (Éd.), *Ingénierie des langues*, Paris: Éditions Hermès, 151-171.

[Abstract]

### **An Epistemological Approach on a Corpus-Based Translation Studies**

Cho, Joon-Hyung  
(Korea University)

*What is the corpus ?* Many researchers have defined this notion in different ways. A *corpus* is the textual collection of linguistic data produced by a language community. In the same way, we can define the *translation corpus* as a collection of translations data produced by professional or non professional translators. The corpus is considered as a computational tool in the field of Natural Language Processing and Machine Translation. And then the main goal of a corpus-based translation studies is to extract translation equivalences from a parallel corpus. But we have to admit that it belongs also to translation studies.

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. The translation studies as human sciences aims at exploring translation phenomena, at analysing them and at explaining the principle of translation phenomena in linguistic, cultural, philosophical views. In this point of view, we have to consider a translation corpus as a language text for a human studies, not as a computational tool, at least in the domain of translation studies. The language text which represent linguistic-cultural translation between different human communities, that is the essential nature of the corpus.

▶ Key Words: corpus, translation corpus, translation studies, epistemological approach, translation, human sciences

조준형

고려대학교 번역과레토릭연구소 연구교수

chojh4net@naver.com

관심분야: 코퍼스 번역학, 번역평가

논문투고일: 2014년 8월 3일

심사완료일: 2014년 8월 29일

게재확정일: 2014년 9월 15일