

## 코퍼스를 활용한 번역 연구 - 국내외 연구 동향과 과제 - \* \*\*

최 희 경  
(한국외대)

### 1. 서론

번역학과 번역 연구에서 ‘코퍼스’라는 술어가 더 이상 낯설지 않은 개념이 된 듯하다. 국내외 주요 학술지에 게재된 논문 중 코퍼스를 실제 구축하고 활용한 논문을 살펴보면 어떤 코퍼스를 어떻게 활용했다는 내용은 있을지언정, 코퍼스가 무엇인지 구구절절 설명하는 논문은 더 이상 찾기 힘들다. 해외 번역 학술지에 실린 코퍼스 관련 논문을 거슬러 올라 검색해보면 코퍼스가 키워드 또는 초록에 나타난 논문은 무려 25년여 전부터 등장한다. 이에 비해 국내 번역 학술지에서는 코퍼스 관련 논문이 등장한 지 불과 15년 정도가 지났을 뿐이

---

\* 이 논문은 2013년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2013SIA5B4A01056140).

\*\* 이 논문은 2015년 번역학회 봄학술대회에서 발표한 내용을 수정, 보완한 것임.

다). 어떻게 보면 국내 번역학계에서는 짧은 기간 동안 코퍼스 연구를 비교적 빨리 도입하고 받아들여 온 셈이다. 이 점에 비추어 볼 때 코퍼스라는 술어에서 느끼는 익숙함의 정도와 코퍼스의 연구 활용도가 반드시 비례하지는 않을 수 있다. 실상 여전히 많은 연구자는 코퍼스 기반 연구를 두려워하거나 난해한 전문분야로 경계하기도 한다(조준형 2014: 289). 본 연구자 또한 코퍼스 연구에 대한 선입견을 가졌던 적이 있다. 코퍼스언어학이라는 분야와 거기에서 파생되는 통계적 분석과 소프트웨어에 압도되고, 무엇이 코퍼스 연구이며 내 연구가 과연 코퍼스 연구가 될 것인지에 대한 의구심을 가졌던 경험이 있다. 코퍼스 연구를 소개한 국내외 다수의 지침서는 코퍼스와 그를 활용한 연구에 대한 기본 지식을 습득할 수 있는 좋은 방법임에는 틀림없지만, 초보자들은 책에서 논하는 연구방법과 자신의 실제 연구 간에 괴리감을 느낄 수도 있다. 본고는 이를 극복하기 위한 방법으로 관심 번역 현상의 분야와 주제에 관계없이 기존의 코퍼스 활용 연구사례를 다수 살펴보고 코퍼스의 개념과 그 방법론이 어떻게 수용되고 있는지를 보여줌으로써, 연구자들이 스스로 코퍼스 활용 여부를 결정하는 데 도움을 주고자 한다. 다시 말해 본고는 코퍼스 연구를 실행할 때 자주 제기되는 주요 논점을 중심으로 실제 연구사례를 살펴봄으로써 교과서적인 이론적 지침을 주기보다 코퍼스 연구에 대한 현실적인 정보를 제공하고 국내 코퍼스 기반 번역 연구의 발전적인 지향점을 찾아보는 데 목적을 둔다.

우선 본고는 코퍼스 기반 번역 연구의 실행에서 고려하게 되는 논점을 정리한 후, 논점별로 국내외 학술지에 실린 연구사례를 개괄 분석한다. 국내외 학술지를 굳이 비교분석하는 이유는 시기상으로 해외에서 코퍼스 기반 번역 연구가 먼저 시작된 데다 해외 학술지가 다수의 국가와 문화를 배경으로 한 연구 특성을 반영하므로 특정 국가나 문화에 치우치지 않은 연구 다양성과 시사점을 기대할 수 있기 때문이다. 또 이를 바탕으로 국내 연구의 현황을 가늠하고 향후 더 발전된 연구를 위한 방향을 모색하기 위해서다. 각 논점별로 국내외 학술지 논문 현황을 살펴본 후에는 국내 연구에 주는 시사점으로 결론을 같음하고자 한다.

---

1) 분석 대상 학술지에 대해서는 3장에 자세히 기술한다.

## 2. 코퍼스 기반 번역 연구-실행상의 논점

베이커(Baker 1993: 235, 248)가 코퍼스언어학의 연구방법을 번역학에 적용해 독립된 연구분야로 발전시킬 것을 주창한 이래, 라비오사(Laviosa 1998, 2002), 올로한(Olohan 2002, 2004), 자네틴(Zanettin 2012) 등 여러 번역학자가 번역학의 맥락에서 코퍼스를 정의하고 그 적용방법과 주요 연구사례를 논의한 바 있다. 코퍼스언어학에서조차 ‘코퍼스란 무엇이다’를 규정하는 완벽한 합의가 이루어지지 않은 상태지만(Saldanha 2009), 번역학 내에서 코퍼스의 개념이 어떻게 수용되었는지는 번역학자들이 채택한 코퍼스의 정의와 속성에서 어느 정도 엿볼 수 있을 것이다(표 1 참조).

〈표 1〉 번역학자들의 코퍼스 정의

학자	코퍼스의 정의	비고
베이커 (1995: 225-26)	1) 원래 (주로 특정 저자의) 가공된 또는 가공되지 않은 글을 모은 것. 2) 코퍼스언어학의 발전으로 새롭게 정의됨. - 자동/반자동 분석이 가능하고 기계 해독이 가능한 텍스트의 모음. - 문어뿐만 아니라 구어도 포괄. - 다양한 출처/저자(화자)/주체의 대규모 텍스트 모음. 3) 구체적 목적 하에 명확한 기준에 따라 구축해 대표성을 확보해야 함.	수작업 연구대상이 되는 소량의 텍스트 모음도 코퍼스의 범위에서 배제하지 않음.
라비오사 (2002: 33)	1) 일반적으로, 명시적인 기준에 따라 수집된 전체 또는 표본 텍스트의 모음으로 정의. 2) 요한손(Johansson 1998: 3)을 인용, “일정 기준에 따라 컴퓨터 처리를 위해 모은 텍스트 덩치”로 정의.	코퍼스언어학에서 사용하는 정의를 채택하나 ST-TT로 구성되는 번역코퍼스의 고유한 특성 지적.
올로한 (2004: 1)	1) 코퍼스란 일정 기준에 따라 선정되고 구축된 텍스트의 모음. 2) 전자적 형태로 되어 있어 각종 코퍼스 소프트웨어로 분석 가능.	코퍼스언어학에서 논의된 내용을 번역학에 적용한다는 관점. 코퍼스는 번역을 다양한 방식으로 연구할 수 있도록 해주는 연구 도구라는 시각.
자네틴 (2012: 7)	1) 일반적으로는 일정한 기준에 따라 수집한 텍스트 모음. 2) 코퍼스언어학에서 코퍼스는 기본적으로 전자화된 텍스트 모음으로, 언어 연구를 위해 고안된 소프트웨어로 가공, 분석되는 대상을 말함.	일반적으로 통용되는 광의의 의미와 코퍼스언어학 맥락에서 사용되는 협의의 의미를 제시.

표에 정리한 정의를 살펴보면 번역학자들은 코퍼스를 코퍼스언어학에서 사용하는 정의로 한정하는 시각뿐만 아니라 ‘일정한 기준으로 수집한 텍스트의 모음’이라는 일반적인 의미로 보는 시각 또한 배제하지 않음을 알 수 있다. 이런 인식은 연구자들의 코퍼스에 대한 수용 태도에도 영향을 줄 수 있다. 코퍼스를 일정 텍스트의 모음으로 본다면 굳이 전자화한 텍스트가 아니라도 자신의 연구 목적에 맞는 텍스트를 모으는 자체로 코퍼스 기반 연구에 한 발 내딛게 되는 것이다. 다만 전자화한 텍스트로는 여러 소프트웨어를 활용한 다양한 정량적인 텍스트 분석이 가능하므로 비전자적인 텍스트에서는 이를 포기하는 대신 다른 분석방법을 설계해야 할 것이다.

표에 나타난 정의에서 공통적으로 언급되는 사항은 바로 코퍼스를 구축하기 위한 명시적이고 구체적인 자료 선정 “기준”이다. 코퍼스를 구축할 때는 자신의 가설과 주장을 정량적으로 검증하거나 정성적으로 설명할 수 있도록 적합한 자료를 선정해야 한다. 대부분의 코퍼스 연구 지침서에서 특히 강조되는 점은 자료의 대표성(representativeness), 균형(balance), 비교가능성(comparability) 등이다. 코퍼스는 결국 자신이 연구하고자 하는 텍스트의 모집단에서 추출한 표본의 집합이므로 표본이 모집단의 속성을 제대로 대표하는지가 연구결과의 신뢰성을 좌우하게 된다. 특히 코퍼스가 여러 장르와 유형을 포괄하는 일반코퍼스(general corpus)인 경우 하위범주의 구성과 비율이 균형 있게 구축되어야 일반적인 텍스트 모집단을 대표할 수 있으므로, 코퍼스의 설계 기준이 매우 치밀해야 하며 이를 논문에서 설득력 있게 기술해야 한다. 코퍼스 설계에서는 코퍼스(또는 각 하위범주)의 크기, 텍스트의 완전성 여부(전체 또는 일부 추출), 생산 및 출판시점, 장르, 유형 등이 따져볼 기준이 된다(자네틴 2012: 47).

그중 코퍼스의 크기는 신뢰성과 밀접한 관련이 있다(ibid.: 42). 코퍼스 구성이 균형적이라 해도 각 하위범주별로 텍스트양이 충분하지 않으면 특히 정량적인 결과를 주장하기에는 무리가 있기 때문이다. 그러나 코퍼스 크기에 대해서는 명시적으로 정해진 기준이 있기보다 연구대상이나 목적에 따라 판단해 설정해야 한다는 의견이 많다. 싱클레어(Sinclair 2004)는 코퍼스의 최소 크기는 연구자의 연구질문과 분석방법, 이 두 가지로 결정할 수 있다고 주장한다. 또 어떤 종류의 코퍼스냐에 따라 기대되는 크기가 달라질 수 있다. 예를 들어 전문코퍼스(specialized corpus)는 특정 현상이 나타날 확률이 높아지므로 상대적으로

크기가 작을 수 있다는 것이다(자네틴 2012: 43). 코퍼스가 작을 경우 연구자가 자신의 주장에 일정 한계와 제약을 미리 설정하고 그 범위 내에서 연구결과를 기술하기도 한다(ibid.: 44). 신뢰성을 인정받은 선행연구를 참고하고 앞서 언급한 고려사항을 충분히 검토한 후 객관성을 갖출 수 있는 코퍼스 크기를 결정하고 설득력 있게 기술하는 일은 결국 연구자의 몫이라 할 수 있다. 단일언어로 된 비교코퍼스 또는 이중(다중)언어코퍼스 비교 시 관심 언어 현상의 차이가 수많은 변수 중에서도 ‘번역 대 비번역의 차이’ 또는 ‘언어 간 차이’로 나타날 수 있도록 통제하기 위해(라비오사 2002: 39), 코퍼스의 비교가능성 또한 반드시 고려해야 한다. 다시 말해 여러 코퍼스를 비교하는 연구에서는 각 코퍼스가 비슷한 주제, 구성, 비율로 구성되었는지 살펴야 한다. 전문코퍼스(specialized corpus)를 비교하는 경우 이는 상대적으로 수월할 수 있다.

코퍼스 기반 번역 연구에서 또 다른 논점은 분석의 접근방법이다. 코퍼스 기반 연구는 통계적 기법과 자동화된 컴퓨터 처리를 활용한 정량적 분석이 그 중심이며, 이를 제대로 이해할 수 없다면 시도하기 힘들다는 인식이 있는 것이 사실이다(조준형 2014: 280). 앞서 정리한 번역학자들의 코퍼스 정의가 전자화된 텍스트와 자동화된 분석에 중점을 두고 있는 것도 이런 인식과 무관하지 않을 것이다. 그런데 이와 관련한 번역학자들의 관점을 보면 이들이 반드시 정량적 분석과 자동화 처리에만 경도된 것으로 보이지는 않는다(올로한 2004: 22). 번역학자들은 코퍼스 연구에 긍정적인 기대를 가지기도 했지만, 코퍼스가 정량적 분석에 치우친 채 누가 봐도 분명해 보이는 가설을 증명하는 데에만 사용되는 것은 경계했다. 숫자 그 이상의 의미를 찾아내거나 숫자에 가려진 진정한 번역의 속성을 찾아내야 한다는 것이다. 이는 곧 분석과 서술에서 정량적인 방법과 정성적인 방법을 결합할 것을 권장하는 것이다(ibid.: 16; Myer 2002: 125). 많은 경우 비교 가능한 코퍼스 간에는 빈도 측정만으로도 흥미로운 결과를 도출할 수 있으며, 빈도 등 정량적 결과는 번역 현상을 더 충실하게 기술하고 원인을 설명하는 정성적 분석을 위한 토대가 된다(올로한 2004: 86).

분석방법과 관련해서 분석의 신뢰성 문제도 제기될 수 있다. 우선 정량적 분석결과에 신뢰성을 부여하기 위해서는 고려할 점들이 있다. 가장 기본적인 양적 분석으로 빈도를 측정했을 때, 서로 다른 크기의 코퍼스 간 비교를 위해서는 빈도를 정규화(normalization)<sup>2)</sup>해야 한다. 또 통계적 분석기법을 적용할 때는

사용하려는 기법이 자신이 고찰하는 연구현상에 적합한지 판단해야 하며, 분석 기법별로 결과의 유의성(significance)을 보여줄 수 있는 지표를 논문에 기술해야 한다<sup>3)</sup>(Biber et al. 1998: 275). 신뢰성 문제는 비단 통계분석 결과의 유의성에만 해당하지 않는다. 예를 들어 번역 현상을 연구자가 평가해 분석하는 경우, 대량의 분석 작업을 여러 명이 분담한다면 평가의 신뢰성 문제가 제기되며, 이 때는 복수 참여자 간 평가일치성(inter-coder reliability)을 검증하고 기술해야 한다(이창수 2013: 159).

그 외에도 번역 연구자는 코퍼스 구축과 관련해 실제적인 제약과 한계에 부딪히게 된다. 예를 들어 인쇄자료를 전자화할 때는 인쇄물을 스캔해서 OCR(문자광학인식) 도구를 사용해 전자문서 형식으로 변환하는 등 상당한 시간, 노력, 비용이 소요되는 번거로운 처리를 거쳐야 한다. 또 자동화된 언어분석을 위해 코퍼스에 언어정보를 삽입, 표시하는 주석처리(tagging) 등 전처리(pre-processing) 작업을 해야 할 때도 있다. 이런 작업은 자동화된 프로그램으로 가능하기도 하지만 대부분 수작업 보정을 거쳐야 하므로 마찬가지로 시간과 노력이 소요된다<sup>4)</sup>. 특히 번역 연구자는 번역-비번역 텍스트로 구성된 단일언어 코퍼스 구축에 비해 출발어텍스트(ST)와 도착어텍스트(TT)로 구성되는 병렬코퍼스 구축에 현실적인 제약을 많이 느끼게 된다. 직접 ST와 TT의 쌍을 대응하면서 자료를 확보해야 하는 데다 전자적인 텍스트로 분석하는 경우 ST와 TT를 문장이나 절 등의 단위로 대응, 정렬(alignment)하는 추가적인 사전 준비가 필요할 수 있기 때문이다. 병렬코퍼스 구축과 관련한 이런 제약을 고려할 때 예산이나 인력이 확보될 수 있는 정부 지원 연구소나 대학 등의 기관에서 연구용

2) 빈도의 정규화(normalization)란 코퍼스 크기가 서로 다른 경우 각각에서 측정된 빈도를 대등하게 비교할 수 없으므로 공통 기준을 정해 상대빈도를 환산해내는 방법을 말한다. 우선 처음 구한 빈도를 원래 코퍼스 크기로 나눈 다음 그 값을 정규화 기준으로 정한 크기에 곱하면 된다(Biber et al. 1998: 263-64).

3) 예를 들어 두 집단 간 차이 분석(t-검정이나 ANOVA 등), 여러 변수 간 상관관계 분석(correlation) 등 분석하려는 현상에 적합한 기법을 선택해야 한다(ibid.: 275).

4) 물론 앞서 코퍼스의 정의에서도 살펴보았듯이, 연구자가 코퍼스를 단순히 “텍스트의 모음”으로 인식한다면 이런 처리와는 거리가 멀 수 있다. 다만 이 경우에는 모든 분석이 수작업으로 이루어질 것이므로 분석 단계에서 시간과 노력이 많이 소모되거나 연구범위가 매우 제한적이 될 수 있다.

코퍼스를 구축해 공개적으로 제공한다면 매우 유용할 것이다. 그러나 국내의 경우 번역 연구자들이 실제 연구에서 활용할 수 있는 기관 구축 코퍼스는 찾기 힘든 실정이다<sup>5)</sup>. 대규모 코퍼스를 구축하기 위한 저작권 문제 또한 코퍼스의 구성과 구축에 실질적인 제약을 주고 궁극적으로 대표성에 영향을 주는 요소다(자네티 2012: 46). 특히 번역 저작권은 확보가 어려우며(베이커 1995: 234), 확보한다 해도 개인 연구자들이 비용을 감당하기는 어렵다. 연구 목적이거나 개인 용도, 교육 용도로 활용할 때 저작권 적용이 예외가 되기도 하지만 그 적용의 기술적 세부사항은 모호한 상태다(올로한 2004: 50-51). 코퍼스 구축의 실행적인 측면에서 나타나는 이런 제약은 코퍼스 기반 번역 연구를 시도하는 번역 연구자들을 위축시키는 요인이 된다.

본 연구는 지금까지 논한 코퍼스 기반 번역 연구의 논점들이 실제 번역 연구에서 어떻게 나타나고 있는지를 살펴봄으로써 코퍼스 기반 번역 연구를 염두에 두고 있는 연구자들이 본고의 내용을 참고해 연구를 시작할 수 있도록 돕는데 목적이 있다. 아래에서는 앞서 논의한 논점을 ‘코퍼스의 정의와 속성,’ ‘분석 방법,’ ‘대표성과 신뢰도,’ ‘코퍼스 구축 실행’의 네 범주로 나누어 각각에 대해 실제 연구를 분석하고 시사점과 향후 과제를 논하도록 한다.

### 3. 국내외 코퍼스 기반 번역 연구의 동향

국내외 코퍼스 기반 번역 연구의 동향을 분석하기 위해서는 분석대상의 범위를 설정할 필요가 있다. 우선 번역학의 범위 내에서 논의된 연구만을 대상으로

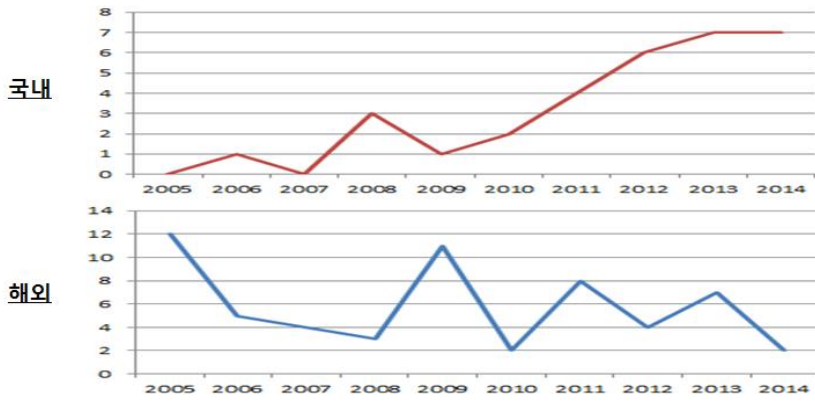
5) 유현경과 황은하(2011)에 따르면 국내에서 구축된 유료 기관 구축 코퍼스로서 세간에 잘 알려진 한영 병렬코퍼스는 세종코퍼스, KORTERM 코퍼스 등이 있으며, 기타 한일, 한중, 한불 코퍼스 등이 연세대학교와 고려대학교를 중심으로 구축된 바 있다. 그러나 본고에서 분석한 국내 번역학 논문을 검토한 결과, 특히 병렬코퍼스의 경우 공개코퍼스를 사용한 사례는 한 건도 없었다. 이는 이들 코퍼스가 개별 연구자들이 연구의 목적에 따라 추출해 활용할 만큼 규모가 크지 않았거나 실제 서비스가 제대로 이루어지지 않았기 때문이다. 세종코퍼스의 경우 개별 신청자에게만 저작권이 해결된 부분에 한해 DVD로 배포되어 규모가 매우 제한적이고, KORTERM의 경우 위 논문에 소개된 서비스 제공 웹사이트 주소가 더 이상 유효하지 않았다.

로 삼았다. 여기에는 번역학 관련 도서, 학위논문, 학술지 논문 등이 포함될 수 있으나, 이 모든 연구를 한꺼번에 정리하기는 현실적으로 불가능하므로 그중 비교적 범위가 분명한 번역학 전문학술지를 대상으로 삼기로 한다. 번역학 전문학술지는 각 학술지에서 표방하는 투고, 심사, 게재방침에 따라 논문의 전개 방향이나 논지가 균질하다고 볼 수 있으며, 번역학 도서나 학위논문에 비해 자료 확보가 용이하다는 장점이 있다. 둘째, 대상 논문의 기간을 한정한다. 해외 논문의 경우 1990년대부터 코퍼스 관련 논문이 검색되는 반면, 국내 논문의 경우 2000년에 처음 코퍼스 기반 번역 연구 논문이 등장한 후 2000년 중반까지 관련 논문이 거의 나타나지 않았다<sup>6)</sup>. 이를 감안해 2005년부터 2014년까지 최근 10년간의 논문을 대상으로 분석하기로 한다. 셋째, 코퍼스 연구를 표방한 연구만을 대상으로 한다. 논문의 제목, 키워드, 초록에서 ‘코퍼스’라는 술어를 사용한 논문을 검색해 추출한다<sup>7)</sup>. 넷째, 코퍼스 연구의 방법론, 연구사, 선행연구 소개 등의 논문은 배제하고 실제로 코퍼스를 구축하거나 구축된 코퍼스를 활용한 논문만을 대상으로 분석한다. 본 연구의 목적이 코퍼스 연구의 실행과 관련한 논점을 실제 연구에서 확인하는 데 있기 때문이다. 이상의 요건을 적용하여 해외 논문으로는 메타(Meta)<sup>8)</sup>에서 58편, 국내 논문으로는 주요 등재지에서 30 편<sup>9)</sup>을 추출했다.

- 6) 메타(Meta)에서는 온라인 검색 결과 1990년 논문에서 이미 코퍼스가 등장한다. 국내 번역학 학술지에서는 학술연구정보서비스(RISS)의 학술지 검색을 활용해 검색한 결과, 2000년, 2003년 각 한 건의 코퍼스 기반 번역 연구 논문이 나타났다.
- 7) 코퍼스 연구면서도 본문 외에 ‘코퍼스’라는 술어를 사용하지 않은 경우에 대비해 데이터 모음, 데이터베이스 등의 키워드로 추가 검색을 실시하였으며, 검출된 논문 각각에 대해 본문 검색을 실행, 누락되는 논문이 없도록 했다.
- 8) 주요 국제 학술지 중 Meta, Perspectives, Target, The Translator, ITT 등을 살펴본 결과, 본 연구에서 정한 요건을 충족하기 위한 검색이 가능한 학술지는 메타가 유일했다. ITT의 경우 2007년부터 간행되어 분석 시기가 맞지 않았고 나머지 세 학술지는 출판사 전체 간행물이 함께 검색되거나, 검색 기능 자체가 제공되지 않아 정교한 검색이 가능하지 않았다. 모든 논문을 일일이 확인하기에는 현실적인 제약이 크고, 메타에서 추출한 논문이 국내 연구의 두 배에 가까운 58편으로 분석에 무리가 없다는 판단 하에 연구범위를 메타로 제한했다. 메타 논문 중에서는 영어로 쓰인 논문만 추렸다.
- 9) 번역학연구, 통번역학연구, 통역과 번역 등 세 학술지의 논문이 포함되었으며, 영어와



그림 1 국내의 학술지 코퍼스 기반 연구 논문의 연도별 추세



대상 논문의 연도별 증감 추세는 위 그래프와 같다(그림 1). 두 그룹의 전체 편수가 달라 연도별 편수를 비교하는 것은 큰 의미가 없지만(그래프상의 척도도 동일하지 않다.), 전체적인 추세를 살펴보는 것은 의미가 있다. 국내 논문은 2005년 전무했으나 최근에 오면서 증가하는 추세를 보이고 있는 데 비해 해외 논문은 2005년에 이미 꽤 많은 편수가 게재되었음을 알 수 있다. 국내의 경우 해외에 비해 코퍼스 기반 번역 연구가 도입된 시기가 10여 년 뒤처져 연구량이 해외의 추세를 뒤늦게 따라가는 양상이라면, 해외의 경우 2005년에 이미 상당한 연구가 진행되었고, 이후에도 뚜렷한 증가나 감소 추세보다는 일정 변동 폭 내에서 증감이 지속되고 있음을 알 수 있다.

## 2.1 코퍼스의 정의와 속성

분석대상 논문에서 구축된 또는 사용된 코퍼스는 전자화되지 않은 단순한 특정 텍스트의 모음에서 정제, 주석처리, 정렬까지 완전하게 이루어진 대규모의 전자텍스트 모음에 이르기까지 모두 나타나 코퍼스에 대한 연구자들의 다양한 인식을 보여준다. 이들 코퍼스의 주요 속성을 정리한 표는 아래와 같다(표 2). 코퍼스의 크기 면에서는 크기를 명시적으로 밝히지 않고 단순히 “큰(또는 작은)” 규모의 코퍼스라고 기술하거나 대사, 문장, 사례, 텍스트의 개수로 나타내

한국어로 쓰인 논문을 모두 포함했다.

는 등 불분명하게 정의하는 경우가 있었다. 이런 경향은 국내와 해외가 별반 다르지 않았다. 국내의 경우 규모를 어절이나 단어로 명시화하지 않은 경우를 제외하고 가장 작은 코퍼스는 52,500 어절이었고 해외의 경우 6,658 단어<sup>10)</sup>였다. 일대일로 비교를 할 문제는 아니지만, 10,000 단어가 되지 않는 텍스트 모음을 코퍼스로 인식하고 통계분석까지 실행한 사실이 특기할 만하다. 코퍼스언어학의 개념을 충실히 수용한 사례를 보면, 국내 연구로는 김혜영(2009)이 200만 어절 규모의 형태소 분석 주석이 달린 번역-비번역 비교코퍼스를 구축, 방대한 정량, 정성 연구를 실행했으며 비번역코퍼스로 공개코퍼스인 세종코퍼스를 활용한 바 있다. 해외 연구에서는 486만 단어 규모의 영-페르시아어 병렬코퍼스 구축 연구(Miangah 2009)에서 연구자가 코퍼스 자체에 내장된 분석 프로그램까지 직접 개발하는 등 코퍼스언어학적 코퍼스 개념을 제대로 구현하고 있다. 국내 연구 중 순수 개인이 구축한 코퍼스 중 최대 규모는 175만 어절이었으며 그중 자동화 분석을 위한 주석달기와 정렬 등 전처리를 거친 병렬코퍼스 사례는 모두 100만 어절 미만의 규모를 보여 코퍼스 구축의 어려움과 한계를 시사한다.

〈표 2〉 코퍼스의 규모 및 속성

구분	국내	해외
크기 (순수 개인 코퍼스)	- 소규모/대규모 - 대사 개수, 작품 개수 - 52,500 ~ 175만 어절	- 소규모/대규모 - 문장 개수, 콘코던스 개수, 텍스트 개수 등 - 6,658 ~ 539만 단어
전처리 (주석/정렬)	30편 중 16편	58편 중 20편
텍스트 유형/주제	연구자가 직접 코퍼스를 구축한 26편 중 20편이 문학 또는 (신문/잡지) 기사	연구자가 직접 코퍼스를 구축한 43편 중 28편이 비문학/비기사 실험 데이터, 광고, UN, 시트콤, 영화, 용어DB 등 다양

10) 카롤리(Károly 2012)에서는 인터넷 뉴스에서 추출한 헝가리어(ST), 영어(TT) 요약기사 각 20개, 6,658 단어의 작은 코퍼스로 원문의 주제구조가 번역에서 어떻게 달라지는지를 고찰했다. 정량적으로는 t-검정과 피어슨 상관관계 분석을 실행했고, 정성적으로는 주제구조가 텍스트 구조와 뉴스 내용에 미치는 영향을 분석했다.

텍스트의 유형과 주제 면에서는 국내와 해외 논문 간에 차이가 나타난다. 물론 일대일 대응이 유의미하지 않을 수도 있다. 해외저널의 특성상 여러 국가의 연구자가 투고를 하는 만큼 본질적으로 연구의 다양성을 내포할 수 있기 때문이다. 그러나 표에서 보듯이 연구가 이루어진 텍스트의 면면을 보면 국내 연구에 시사점을 주는 것도 사실이다. 국내 연구는 코퍼스 자료가 문학이나 기사에 치우쳐 있는 반면 해외 연구는 그 외 다양한 주제나 유형에서 코퍼스 연구가 이루어지고 있음을 알 수 있다. 이 결과는 자신의 연구를 미처 코퍼스 연구와 연계하지 못했던 많은 연구자가 코퍼스 방법론의 적용 가능성을 재고해볼 여지가 있음을 보여준다.

## 2.2 코퍼스 분석방법

코퍼스 분석방법에서는 양극단의 사례가 모두 나타난다. 한편으로는 수작업 분석만 나타나고 다른 한편으로는 고도의 자동화 분석과 통계분석이 나타난다. 정량적, 정성적 분석의 측면에서도 둘 중 하나의 방법만 사용하는 연구에서 둘 다를 활용하는 연구까지 다양하게 나타난다. 국내 연구의 경우 정량적 분석만 실행한 논문은 30편 중 7편으로 주로 통계기법을 사용해 분석을 전개한다. 코퍼스에 정성적 분석만 실행한 연구는 4편이었다. 해외의 경우 코퍼스 활용을 표방하면서도 정성적 분석에만 의존한 경우가 상대적으로 많아 58편 중 32편에 이르렀다. 정량적, 정성적 연구를 모두 실행한 연구는 국내 19편, 해외 20편으로, 표본의 크기가 다르다는 점을 감안하더라도 전체적으로는 국내 연구보다 해외 연구에서 정성적 분석이 중시되는 경향을 보인다. 바꿔 말하면 국내에서는 코퍼스 기반 번역 연구가 정량적 분석 중심이라고 인식하는 정도가 해외에 비해 상대적으로 높다고 할 수 있다.

정량적 분석의 기법을 살펴보면 대부분 빈도와 비율을 분석하고 있으며 이는 국내외 연구에서 별반 차이가 없어 보인다. 그런 가운데 국내 연구에서는 소수의 특정 연구자에 의한 적극적인 통계 활용이 눈에 띈다. 이들의 논문이 분석 대상 논문의 1/3 이상을 차지하고 있으며, 나머지 논문에서는 간단한 통계검정이나 빈도, 비율로 정량분석을 실행하고 있다. 정성적 분석에서 가장 눈에 띄는 방식은 예문을 나열하고, 각각에서 관심 번역 현상이 어떻게 나타나고 있는지

기술하는 형태다. 국내외 연구에서 모두 이런 방식이 주로 사용되고 있으며 특히 양적결과(이러테면 빈도나 비율)를 제시한 후 예문으로 보강하는 전개가 많이 이루어지고 있다. 그러나 양적 분석결과를 바탕으로 기술하는 분석 외에 그 현상이 나타나게 된 원인, 배경, 요인 등을 설명하는 분석은 국내외 모두 뚜렷이 나타나지 않았다. 코퍼스 기반 번역 연구가 기술번역학(descriptive translation studies, DTS)을 토대로 했다는 점(라비오사 2002: 5), 그리고 DTS가 번역물, 번역물의 사회맥락적 기능, 번역가의 인지과정 등 세 관점으로 번역을 연구한다(Holmes 2000: 176-77)는 점에서 번역물 외 번역환경이나 번역가의 의사결정 등 번역 결과물에 영향을 주는 요인을 더 심층적으로 분석할 필요가 있다고 본다.

마지막으로 분석 소프트웨어 사용 측면을 보면, 국내외 공히 가장 많이 사용된<sup>11)</sup> 소프트웨어는 워드스미스(WordSmith), 앤티콘크(AntConc) 등이었고, 해외에서는 이에 더해 코퍼스워크벤치(CWB)도 다수 연구에서 사용되었다. 이 밖에도 해외에서는 전자화된 대규모 기관 코퍼스 자체에 내장된 분석도구들이 자주 사용된 점이 국내 연구와의 차이점이다.

〈표 3〉 코퍼스 분석방법 및 분석도구

구분	국내	해외
정량적 분석	<ul style="list-style-type: none"> <li>- 대부분 빈도/비율 분석</li> <li>- 코퍼스 분석도구를 활용한 기본 통계량 제시</li> <li>- 특정 연구자를 중심으로 통계기법 적극 활용</li> </ul>	<ul style="list-style-type: none"> <li>- 대부분 빈도/비율 분석</li> <li>- 일부 간단한 통계검정 사용</li> </ul>
정성적 분석	<ul style="list-style-type: none"> <li>- 대부분 관심 현상에 대한 예문 분석</li> <li>- 양적결과에 대한 설명적 원인 분석은 상대적으로 적음</li> </ul>	<ul style="list-style-type: none"> <li>- 대부분 관심 현상에 대한 예문 분석 중심이나 상대적으로 세분화된 분석이 관찰됨</li> <li>- 양적결과에 대한 설명적 원인 분석은 상대적으로 적음</li> </ul>
SW 활용	워드스미스, 앤티콘크	워드스미스, 앤티콘크, 코퍼스워크벤치

11) 세 편 이상에서 사용된 소프트웨어만 추렸다.

### 2.3 대표성과 신뢰도

코퍼스 구축 및 분석에서 가장 중요하다고 볼 수 있는 대표성과 신뢰도를 살펴본 결과는 표 4에 정리했다. 우선 코퍼스 구성 및 자료 선정기준을 기술했는지를 확인한 결과, 국내 연구의 경우 상세한 기술이 많지 않았다. 어떤 자료를 선정했고, 양이 어느 정도인지를 기술한 정도에 그친 논문이 많았다. 다만 코퍼스 연구를 지속적으로 해온 특정 연구자들을 중심으로 자료 선정 및 코퍼스 구성에 대한 배경과 근거가 상세히 기술되었다. 해외 연구의 경우 대체로 어떤 틀을 가지고 코퍼스를 구축하는지와 자료의 선정 배경, 종류, 분량, 기간, 주제 등을 비교적 상세히 기술하고 있고, 비교코퍼스를 구축한 논문의 경우 비교되는 코퍼스가 얼마나 균질하고 비교 가능한지를 언급하기도 했다. 국내 연구에서 관련 기술이 상대적으로 적은 이유는 앞서 살펴본 코퍼스 속성과도 관련이 있을 수 있다. 코퍼스를 구성하는 텍스트의 유형과 주제가 문학작품이나 기사 위주이기 때문에 일종의 전문코퍼스로 볼 수 있는 코퍼스가 많고, 따라서 코퍼스의 속성이 직관적이고 균질하다고 생각될 수 있다. 하지만 코퍼스 연구에서 대표성과 신뢰도가 논점이 되는 만큼 연구자가 코퍼스의 종류와 특성에 따라 자료 선정과 구축 기준을 상세히 기술한다면 연구의 대표성과 신뢰도를 확보하는 최소한의 장치를 마련하는 셈이다.

기타 구체적인 분석의 신뢰도 면에서는 통계분석의 경우 국내외 연구 모두 분석의 유의성(significance)을 증명하는 결과를 함께 보여줌으로써 통계분석 신뢰도를 확보하고자 했다. 그런데 국내외 연구가 대체로 특정 통계기법을 사용하는 이유에 대한 명확한 설명이 없었던 공통점이 있었다. 자료의 특성과 연구 질문에 적합한 통계기법을 선정하고 그 과정을 설명한다면 논문의 독자에게 더 많은 정보를 주고 신뢰도도 높일 수 있을 것이다. 한편 해외 연구에서는 복수 평가자, 복수 코더 간 신뢰도를 언급한 논문이 두 편 있었던 데 비해 국내 연구에서는 없었다. 물론 소논문에 언급되지 않는 많은 사항이 학위논문이나 도사에서 언급되기도 하지만, 연구자들이 쉽게 접할 수 있는 참고문헌에서 코퍼스 연구의 논점들에 대한 더 상세한 기술이 이루어진다면 연구의 신뢰도도 높아질 뿐 아니라 독자에게 더 유용한 정보를 전달하게 될 것이다.

〈표 4〉 코퍼스의 대표성과 분석 신뢰도

구분	국내	해외
코퍼스 구성 및 자료 선정 기준	- 상대적으로 자료 선정 및 구성에 대한 기술이 간략하거나 부재함 - 특정 연구자를 중심으로 상세한 기술이 관찰됨	- 코퍼스 구축을 위한 기본틀(frame) 기술 - 자료 선정배경, 종류, 분량, 기간, 주제 등을 상세히 기술 - 비교코퍼스 구축 시 비교가능성(comparability) 언급
분석 신뢰도	- 통계분석의 경우 통계적 유의성 기술	- 통계분석의 경우 통계적 유의성 기술 - 복수 평가자/코더 간 신뢰도 언급(2편)

## 2.4 코퍼스 구축 실행

코퍼스 구축의 실행적 측면에서는 우선 연구자 개인이 구축한 코퍼스를 활용한 사례와 연구자가 개인 구축 코퍼스와 기관 코퍼스를 함께 활용한 사례, 마지막으로 연구자가 기관 코퍼스만 사용한 사례를 확인해보았다. 빈도로만 보면 국내 논문과 해외 논문 간 편수 차를 고려할 때 비슷한 비율을 보인다. 다만 국내 연구의 경우 특정 연구자를 중심으로 기관 코퍼스를 사용하거나 자신의 개인 구축 코퍼스를 기관 코퍼스와 함께 사용한 사례가 다수라 일반 연구자가 기관 코퍼스를 충분히 활용하고 있다고 보기는 어려웠다. 해외 논문의 경우 다양한 국적의 논문이 모여 있으므로 한 국가의 사정은 알기 어렵지만, 국내 논문에 비해 기관이 구축한 병렬코퍼스<sup>12)</sup>를 잘 활용한 사례들이 눈에 띈다.

〈표 5〉 코퍼스 구축 주체 및 기관 코퍼스 활용

구분	국내	해외
연구자 vs. 연구자+기관 vs. 기관	- 연구자(20) vs. 연구자+기관(6) vs. 기관(4) - 특정 연구자 중심으로 기관 코퍼스 사용	- 연구자(41) vs. 연구자+기관(2) vs. 기관(15)

12) 영어-포르투갈어 병렬코퍼스 콤파라(COMPARA) (Saldanha 2012), 영어-네덜란드어 병렬코퍼스 디피씨(Dutch Parallel Corpus, DPC) (Vanderbauwhede et al. 2011), 영어-카탈로니아어 병렬코퍼스 코발트 (COVALT) (Marco 2009), 영어-스페인어 병렬코퍼스 악트레스(ACTRES) (Izquierdo 2014) 등.

기관(공개) 코퍼스 속성	- 균형코퍼스 - 비교코퍼스 위주 활용	- 균형코퍼스, 전문코퍼스 - 비교/병렬코퍼스 두루 활용
공통	- 연구자 개인이 병렬코퍼스 구축 시 코퍼스 크기 제한적 - 저작권 언급이 거의 없음(해외 논문 3편: 저작권 사용 승인된 기관 코퍼스 2편, 저작권 문제 언급 1편)	

국내 논문에서 사용한 기관 코퍼스는 주로 세종코퍼스에서 추출한 한국어 비번역코퍼스로 연구자의 한국어 번역코퍼스와 함께 비교코퍼스 구축에 사용되었다. 또 대학에서 문학 번역코퍼스를 구축해 비공개로 사용한 사례가 관찰되었다. 해외 연구에서는 논문에 기술된 기관 코퍼스 대부분을 온라인 검색으로 쉽게 찾을 수 있었고, 이들 코퍼스가 일반에 공개되었을 뿐 아니라 지속적으로 추가, 보완되고 있음을 알 수 있었다. 이런 연구 인프라는 일반 연구자가 코퍼스 구축과 실행의 부담을 덜고 자신의 관심 현상을 분석하는 데 집중할 수 있도록 해준다. 병렬코퍼스뿐만 아니라 영화<sup>13)</sup>, 해양공학(coastal engineering)<sup>14)</sup> 등 전문코퍼스 또한 기관 주도로 구축, 활용되고 있는 점이 특기할 만하다.

국내외 연구에서 공통적으로 관찰된 점은 연구자 개인이 병렬코퍼스를 구축했을 때 코퍼스 크기가 작아진다는 점으로 이는 앞서 코퍼스 속성을 논할 때 살펴본 바 있다. 또한 국내외 공히 코퍼스 내 텍스트의 저작권 사용에 관한 언급이 거의 없었다. 해외 논문에서 저작권 사용 승인을 받아 코퍼스를 구축한 사례가 소개되었으나 이는 기관 코퍼스 사례였고, 저작권 문제를 언급한 다른 논문은 학생번역의 사례로 특수한 경우였다. 저작권을 득했으면서도 언급하지 않은 사례도 물론 있겠지만 개인 연구자들의 경우 “연구 용도”에는 저작권 적용 예외(올로한 2004: 50-51)라는 인식에 기대고 있는 것으로 보인다.

13) 파비아 영화대사 코퍼스(Pavia corpus of film dialogue) (Pavesi 2013)

14) 푸에르토 용어 코퍼스 (Puerto Term Corpus) (Benítez et al. 2005)

### 3. 향후 과제 및 결론

본고에서는 코퍼스 기반 번역 연구의 실행적 논점을 정리하고 이를 바탕으로 국내외 주요 학술지에 실린 관련 논문의 동향을 검토, 확인하였다. 애초에 본 연구의 목적은 수많은 문헌에서 설파하고 있는 코퍼스 기반 번역 연구가 실제 연구자들에게 어떻게 수용되고 있는지를 살펴 코퍼스 연구를 시작하려는 국내 연구자들에게 도움을 주려 한 데 있으므로, 본 연구자가 작금의 연구 동향에 대해 어떤 가치평가를 내리는 일은 적합하지 않다. 다만 국내에서 코퍼스 기반 번역 연구 관련 논문이 증가하는 추세를 볼 때 더 많은 연구자가 관심을 가지고 연구를 시도하게 될 것은 분명해보이므로, 본고의 고찰에서 얻은 시사점을 나눈다면 그들의 시행착오를 줄이고 연구 가능성을 확대하는 데 미약하나마 도움이 될 수 있을 것이다.

우선 코퍼스 기반 번역 연구의 대상이 더 다각화된다면 코퍼스 기반 번역 연구의 저변이 훨씬 확대될 것이다. 앞서 살펴보았듯이 다양한 텍스트 유형과 주제를 볼 수 있었던 해외 연구와 달리 국내 연구는 문학 또는 기사 위주로 코퍼스가 구축되고 있다. 코퍼스 연구에 관심을 가진 연구자는 코퍼스 개념의 수용 양상과 적용 분야가 실제 연구에서 다양하게 나타난다는 점을 인지하고, 자신의 관심분야에 적합한 형태의 코퍼스 연구를 시도해볼 수 있을 것이다. 또한 번역 연구 자체가 멀티미디어 등 새로운 텍스트 영역으로 확장되고 있는 만큼 이런 영역에서도 코퍼스 기반 접근방식을 차용한다면 연구의 다양성이 곱절로 확대될 수 있을 것이다. 특히 전문코퍼스는 일반코퍼스에 비해 크기도 상대적으로 작을 뿐더러 대표성에서도 높은 수준을 달성할 수 있으므로(자네틴 2012: 47), 개별 연구자들이 각자의 관심 영역에서 처음 코퍼스 연구로 시도하기에 상대적으로 수월할 수 있다. 본고에서는 번역 연구만 대상으로 했지만 통역에도 코퍼스를 구축해 연구하는 사례가 많아지고 있으므로 이런 동향도 참조할 만하다(Kruger et al. 2011 참조). 단 어떤 코퍼스를 구축하더라도 자료 선정기준과 코퍼스 구성을 세밀하게 설계해야 한다는 점은 유념해야 한다.

둘째, 분석방법에서 정량적 분석과 정성적 분석 간 균형을 이룰 필요가 있다. 국내 연구의 분석방법을 수치상으로만 보면 정량적, 정성적 분석을 모두 실행한 논문이 30편 중 19편으로 높은 비율을 차지해 대체로 두 분석방법 간에



균형을 이룬 듯하다. 그러나 질적분석이 양적 분석결과를 그대로 기술하거나 예문을 나열하는 데 그친 경우가 많았다. 코퍼스 분석이 곧 양적분석이라는 등식에 매몰되지 않고 비코퍼스 연구에서 사용하는 심화된 정성분석을 적용하면서 양적결과의 요인을 찾고 설명하는 분석을 전개할 필요가 있다. 정량적 분석에서도 분석을 위한 분석이 되지 않도록 경계할 필요가 있다. 통계기법을 사용하더라도 기초적인 빈도가 주는 정보 그 이상을 주지 못할 때가 많다. 따라서 특정 정량적 기법을 왜 사용하는지, 그 기법이 왜 적합한지를 기술하고 목적에 따라 꼭 필요한 기법을 사용하는 것이 좋다.

셋째, 국내 코퍼스 기반 번역 연구는 그 역사가 짧음에도 해외 연구와 비교 시 논점별로 확연한 차이를 보이지는 않았다. 그런데 연구자들의 코퍼스 구축 실행 측면에 크게 도움이 될 수 있는 기관(공개) 코퍼스의 구축 및 보급 부분에서는 비교적 차이가 뚜렷했다. 국내에는 정부 지원 사업으로 구축된 세종코퍼스가 있기는 하지만 프로젝트 완료 후 더 이상 텍스트가 추가되지 않는 실정이다<sup>15)</sup>. 한영 병렬 말뭉치의 경우 균형코퍼스라 하기에는 포함된 텍스트의 종류가 제한적이고 그마저도 한-영, 영-한 두 방향의 하위코퍼스를 모두 합해 156만여 어절에 그쳐 연구자들이 연구 목적에 맞게 표본을 추출해 사용하기 힘든 상황이다. 사실상 유일한 공개 코퍼스로서의 세종코퍼스는 존재만으로 감사한 일이다. 그러나 번역 연구의 토대가 되는 코퍼스를 정부, 대학, 연구기관 등에서 추가로 구축하고 개인 연구자들에게 제공해준다면 코퍼스 기반 번역 연구의 문턱이 낮아지고 관련 연구가 한층 활성화될 수 있을 것이다. 연구자들은 방대한 시간, 비용, 노력을 들여 대표성 있는 코퍼스를 구축해야 하는 부담을 덜고 분석방법과 연구 신뢰도를 높이는 데 더 전념할 수 있을 것이다.

본고에서는 코퍼스 기반 번역 연구의 주요 논점을 기준으로 국내외 학술지에 실린 코퍼스 기반 번역 연구의 전체적인 그림을 보여주는 데 중점을 두었기에, 개별 논문들의 구체적인 주제별 논지 전개나 분석방법을 자세히 기술하거나 범주화해 제시하지는 못한 한계가 있다. 그러나 논문 검토 과정에서 비슷한 언어 현상이나 번역 현상이 언어나 텍스트를 달리하거나 분석방법을 달리해 반

15) 국립국어원 언어정보나눔터에서 제공하는 한영 병렬 말뭉치 목록을 보면 가장 최근 자료가 2005년 자료로 10년 전이다. 최근인 2014년에 등록된 파일도 세종코퍼스 구축 당시 만들어진 자료로 확인된다.

복적으로 연구되는 경우가 있음을 알 수 있었다. 이런 연구를 범주화해 각각을 더 자세히 살펴보는 일은 본 연구자뿐만 아니라 코퍼스 기반 연구를 실행하려는 동료 연구자들에게도 도움이 될 수 있으리라 생각하며, 이를 후속연구 과제로 기약하고자 한다.

### 참고문헌

- 국립국어원 언어정보나눔터 (2014) 「말뭉치목록\_한영병렬」. 2015년 3월 22일 검색. <https://ithub.korean.go.kr/user/total/referenceView.do>.
- 김혜영 (2009) 『국어 번역 글쓰기의 연구』, 고려대학교 대학원 박사학위논문.
- 유현경, 황은하 (2011) 「병렬말뭉치 구축과 응용」, 『언어사실과 관점』25: 5-40.
- 이창수 (2013) 「코퍼스 기반 번역연구(CTS) 방법론」, 정호정 (편저), 『통번역학 연구 현황과 향후 전망 I』, 서울: 한국문화사, 125-67.
- 조준형 (2014) 「코퍼스 기반 번역 연구의 인식론적 고찰」, 『번역학연구』 15(3): 279-301.
- Baker, Mona (1993) ‘Corpus Linguistics and Translation Studies: Implications and Applications’, in Mona Baker, Gill Francis and Elena Tognini-Bonelli (eds) *Text and Technology: In Honour of John Sinclair*, Amsterdam & Philadelphia: John Benjamins, 233-50.
- Baker, Mona (1995) ‘Corpora in Translation Studies: An Overview and Some Suggestions for Future Research’, *Target* 7(2): 223-43.
- Benítez, Faber Benítez, Linares, Carlos Márquez and Expósito, Miguel Vega (2005) ‘Framing Terminology: A Process-Oriented Approach’, *Meta* 50(4).
- Biber, Douglas, Conrad, Susan and Reppen, Reppen (1998) *Corpus Linguistics: Investigating Language Structure and Use*, Cambridge: Cambridge U. P.
- Holmes, James S. (2000) ‘The Name and Nature of Translation Studies’, in Lawrence Venuti (ed) *The Translation Studies Reader*, London & New York: Routledge, 172-85.

- Izquierdo, Marlén (2014) 'On Describing Similarity and Measuring Equivalence in English-Spanish Translation', *Meta* 59(1): 140-59.
- Johansson, Stig (1998) 'On the Role of Corpora in Cross-linguistic Research', in Stig Johansson & Signe Oksefjell (eds) 3-24.
- Károly, Krisztina (2012) 'News Discourse in Translation: Topical Structure and News Content in the Analytical News Article', *Meta* 57(4): 884-908.
- Kruger, Alet, Wallmach, Kim and Munday, Jeremy (eds) (2011) *Corpus-based Translation Studies: Research and Applications*, London/New York: Continuum.
- Laviosa, Sara (1998) 'The Corpus-based Approach: A New Paradigm in Translation Studies', *Meta* 43(4): 474-79.
- Laviosa, Sara (2002) *Corpus-based Translation Studies: Theory, Findings, Applications*, Amsterdam: Rodopi.
- Marco, Josep (2009) 'Normalisation and the Translation of Phraseology in the COVALT Corpus', *Meta* 54(4): 842-56.
- Meyer, Charles Frederick (2002) *English Corpus Linguistics: An Introduction*, Cambridge: Cambridge U. P.
- Miangah, Tayebeh Mosavi (2009) 'Constructing a Large-Scale English-Persian Parallel Corpus', *Meta* 54(1): 181-88.
- Olohan, Maeve (2002) 'Corpus Linguistics and Translation Studies: Interaction and Reaction', *Linguistica Antverpiensia* 1: 419-29.
- Olohan, Maeve (2004) *Introducing Corpora in Translation Studies*, New York: Routledge.
- Pavesi, Maria (2013) 'This and That in the Language of Film Dubbing: A Corpus-Based Analysis', *Meta* 58(1): 103-33.
- Saldanha, Gabriela (2009) 'Principles of Corpus Linguistics and Their Application to Translation Studies Research', *Tradum tica*, March 3. Available at <http://webs2002.uab.es/tradumatica/revista/num7/articles/01/art.htm>.
- Saldanha, Gabriela (2012) 'Emphatic Italics in English Translations: Stylistic Failure or Motivated Stylistic Resources?', *Meta* 56(2): 424-42.

- Sinclair, John (2004) 'Developing Linguistic Corpora: A Guide to Good Practice', March 22. Available at [www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm](http://www.ahds.ac.uk/creating/guides/linguistic-corpora/chapter1.htm).
- Vanderbauwhede, Gudrun, Desmet, Piet and Lauwers, Peter (2011) 'The Shifting of the Demonstrative Determiner in French and Dutch in Parallel Corpora: From Translation Mechanisms to Structural Differences', *Meta* 56(2): 443-64.
- Zanettin, Federico (2012) *Translation-driven Corpora*, Manchester & Kinderhook: St. Jerome.

[Abstract]

**The Current Status and Trends of Corpus-based Translation Studies**

Choi, Hee-kyung

(Hankuk University of Foreign Studies)

Since Baker (1993) first advocated the use of corpora in translation studies, the corpus-based translation studies (CTS) field has gradually broadened its presence in the broader field of translation studies. Today, the term “corpus” has become quite familiar among translation researchers. However, such familiarity ironically makes them reluctant to initiate their own corpus-based research as they have become more conscious of presumed technical difficulties associated with building and analyzing a corpus. Indeed, CTS is widely known as a methodology that is good for tech-savvy researchers. Against this backdrop, the present study aims at identifying the current status and trends of CTS at home and abroad, in an attempt to shed light on “CTS in reality.” To this end, this study reviews 88 CTS-related journal articles—30 domestic and 58 international—in light of arguably the most debated topics in CTS: the definition of a corpus; methods of corpus analysis; representativeness and reliability; and practical issues with building and using a corpus. Findings reveal that the reviewed articles show a whole spectrum of CTS varieties, which the author hopes may help potential CTS researchers find ways to easily start their own studies with more confidence.

▶ Key Words: corpus-based translation studies, corpus design, quantitative and qualitative analysis, representativeness, reliability

최희경

한국의국어대학교 영어대학 EICC학과

trans.hufs@gmail.com

관심분야: 코퍼스 기반 번역 연구, 언어접촉 및 변화, 번역교육

논문투고일: 2015년 5월 3일

심사완료일: 2015년 5월 31일

게재확정일: 2015년 6월 8일