

코퍼스기반 번역학 연구에서 정량적 인자가 정성적 분석 결과에 미치는 영향*

김 정 우
(경남대)

1. 머리말

코퍼스(말뭉치) 번역학이 갖는 이론적 장점은 자명하다. 그것은 자료의 규모가 자료 자체에 내재한 어떤 경향성을 비교적 선명하게 통계 수치로 드러낼 수 있다는 사실이다. 그리하여 방대한 분량의 실제 번역 자료를 활용한 번역학 연구자들이 번역 보편소의 존재를 가설에서 이론으로 격상시킬 수 있었던 것도 코퍼스 번역학의 성과임은 부인할 수 없다. 그렇지만 코퍼스 번역학은 장점 못지않게 한계 또한 분명한 연구방법론이다.

* 이 논문은 2010년도 정부재원(교육과학기술부 인문사회연구역량강화사업비)으로 한국연구재단의 지원을 받아 연구되었음(NRF-2010-327-A00582). 과제 제목은 ‘한국어 번역문의 중간언어적 특성과 그 극복 -정량적 코퍼스 분석에 기초한 연구 성과의 실용성 극대화를 위한 정성적 연구-’인데, 본고는 2차 연도 연구의 결과물이다.

코퍼스 번역학의 원론적 한계는 대규모 코퍼스 분석을 통해 얻어진 통계 수치가 기본적으로 결과론적 기술이라는 데 있다. 즉, 대규모 코퍼스의 통계 분석 결과는 원전이나 번역에 나타난 특징을 계량화해서 보여줄 뿐, 원전과 번역의 상관관계 혹은 다양한 요인이 경합하여 역동적으로 진행되는 번역 과정의 해명에는 미치지 못한다.¹⁾ 그러므로 대규모 코퍼스를 활용한 정량적 통계 처리의 결과물을 실제 번역 작업과 번역 교육에 기여할 수 있는 번역 매뉴얼로 형식화하기 위해서는, 통계 수치를 액면 그대로 수용하기보다, 번역과 원전 자료의 비교 및 대조에 기반한 정성적 분석을 통해 통계 수치의 이면에 도사린 의미를 읽어낼 수 있도록 연구자의 능동적인 보완 작업이 필요하다는 것이다.

이러한 맥락에서 우리는 영어 원문과 한국어 번역문으로 구성된 일정한 규모의 병렬코퍼스를 활용하여 영-한 번역문의 특징이 어떠한 번역 과정에서 비롯된 것인가를 분석한 바 있었다(김정우 2011). 그런데 여기서 한 가지 의문이 자연스럽게 대두된다. 그것은 우리의 분석이 신뢰할 만한 결과가 되려면, 과연 코퍼스의 규모를 어느 정도까지 늘려야 하는가의 문제이다. 물론 자료의 규모가 크면 클수록 그로부터 도출된 관찰이나 분석의 결과도 높은 일반성을 확보할 수 있고, 자료의 경향성이 간직한 ‘사실’에 근접할 가능성이 높아질 것이다. 그러나 통상적인 연구 환경에서 자료의 규모를 무한정 늘리기란 어려운 것이 사실이므로, 분석 결과의 일반성을 담보할 수 있는 코퍼스의 적정 규모를 제안할 수 있다면 코퍼스 번역학의 이론적 발달에 의미 있는 기여를 할 것으로 생각된다.

본고는 병렬코퍼스를 활용한 정성적 분석 결과에 코퍼스의 양적 규모가 미치는 영향 관계를 실증적으로 규명해보려는 시도이다. 이러한 목표를 달성하기 위해 우리는 코퍼스의 규모를 일정한 크기로 누적시켜 가면서 분석의 결과가 어떠한 변화를 보이는가를 관찰할 것이다.

본고의 구성은 다음과 같다. 제2장에서는 25만 어절의 병렬코퍼스를 대상으로 한 정성적 분석 결과를 요약하여 이어지는 논의의 출발점으로 삼는다. 제3

1) 자연언어처리 기술 수준이 지금보다 발달하여 더욱 정밀한 연구 설계가 이루어진다면, 원문과 이에 상응하는 번역문의 상관관계가 프로그램상에서 자동적으로 추출될 가능성도 없지는 않다. 그러나 선택적 명시화의 사례에서 보듯이 현재의 연구 여건에서 프로그램이 행간의 의미까지 읽어내기란 아직도 요원한 경지로 보인다.

장과 제4장에서는 양적 인자의 증가에 따라 정성적 분석의 결과가 어떻게 변화하는가를 관찰할 것이다. 우리는 15만 어절씩 자료를 추가할 때마다 발생하는 분석 결과의 변화 추이를 면밀하게 비교하면서, 양적인 인자의 누적된 변화가 어떻게 분석 결과에 질적으로 의미 있는 영향을 미쳤는가를 조사하여, 분석의 일반성을 담보할 수 있는 적정 코퍼스의 규모를 확정하게 된다. 제5장에서는 본고에서 논의한 내용을 요약하면서 글을 마무리하게 된다.

2. 25만 어절 병렬코퍼스 분석 결과

여기서는 영어 원문과 한국어 번역문으로 이루어진 25만 어절의 병렬코퍼스를 활용한 원문과 번역문의 대조 분석 결과를 요약하여, 앞으로 이루어지는 분석의 대상과 내용을 구체적으로 밝히고 이를 후속 논의의 출발점으로 삼고자 한다.

김정우(2011)에서 관심을 가진 영-한 번역문의 언어 층위별 조사 대상 항목은 어휘와 형태 층위에서 의존명사 ‘때문’의 고빈도 출현, 구문 층위에서 ‘-게 하다’와 ‘-게 만들다’로 대표되는 장형 사동문의 우세적 출현, 담화 층위에서 문장 접속부사 ‘그러나’와 ‘그런데’ 및 ‘하지만’ 등의 상대 분포, 문장부호 차원에서 줄표(dash)의 번역 양상 등 네 가지 특징이었으며, 우리는 이들 네 가지 특징에 초점을 맞추어 영어 원문과 한국어 번역문을 비교·검토하여 이들 특징과 관련된 몇 가지 사실을 밝혀보았다. 결과를 항목별로 정리해보면 다음과 같다.

2.1. 어휘·형태 층위 - 의존명사 ‘때문’

병렬코퍼스를 활용한 의존명사 ‘때문’의 용례 조사를 통해서 확인하고자 한 내용은 다음의 두 가지였다. 첫째는 의존명사 ‘때문’의 출현 분포를 조사해보았다. 이러한 조사를 통해서, 우리는 번역문에서 의존명사 ‘때문’이 상대적으로 높은 분포를 보이는 원인을 ‘-기 때문’ 구성의 빈출에서 찾고 이를 어휘적 단순화와 관련시킨 김혜영(2009a: 42-45)의 주장과 대조해보았다. 둘째는 번역 중재 과정을 해명하기 위해 의존명사 ‘때문’이 원문의 무엇을 번역한 결과인지를 알

아보았다.

우리가 구축한 한국어 번역문 코퍼스에서 의존명사 ‘때문’이 포함된 문장 용례는 모두 559건이었다. 분포 조사의 결과는 다음과 같이 네 가지로 나타났다. 의존명사 ‘때문’은 ‘-기 때문’ 구성으로 가장 많이 나타났는데(417회, 74.6%), 이는 김혜영(2009: 42-45)의 지적과 대체로 일치했다. 다음으로 ‘명사+대명사+때문’ 구성(131회, 23.4%), ‘관형사+때문’ 구성(10회, 1.8%) 및 단독형 ‘때문’(1회, 0.2%)의 순서로 나타났다.

원문과의 대조를 통해 의존명사 ‘때문’의 번역 연원을 조사한 결과는 다음과 같았다. 대체로 접속사를 번역했을 것이라는 예상과 달리, 실제로는 전후 문맥에서 유추해서 영형태(零形態, zero)를 ‘때문’으로 번역한 용례가 가장 많았다(249회, 44.5%). 이보다 조금 낮은 빈도로 나타난 유형은 접속사를 번역한 용례였다(223회, 39.9%). 이어서 전치사를 번역한 용례(40회, 7.2%), 특정 구문을 번역한 용례(19회, 3.4%)와 문장부호 줄표와 반점을 번역한 용례(17회, 3.0%)가 엇비슷한 빈도로 나타났으며, 명사나 부사를 번역한 용례(11회, 2.0%)가 가장 적었다. 의존명사 ‘때문’의 번역 연원 조사에서 원문에 없지만 전후 문맥에서 유추해서 번역문에 도입한 용례가 가장 많았다는 사실은 번역 보편소 가운데 하나인 선택적 명시화의 특징이 드러난 전형적 사례로 볼 수 있다.

2.2. 구문 층위 - 장형사동

병렬코퍼스를 활용한 장형사동 구문 ‘-게 만들다’/‘-게 하다’의 용례 조사를 통해서 확인하고자 한 내용은 다음의 두 가지였다. 첫째는 추출된 장형사동 구문의 용례 가운데서 접미사의 첨가에 의한 단형사동 구문으로 교체할 수 있는 용례가 얼마나 되는지를 알아보았는데, 이는 번역문에 통사적 사동, 즉 장형사동이 우세하게 나타난다는 기존 논의(김혜영 2009a: 108-110)의 검증 작업이었다. 둘째는 번역 중재 과정을 해명하기 위해 장형사동 구문이 원문의 어떤 표현을 번역한 것인지를 알아보았다.

우리가 구축한 한국어 번역문 코퍼스에서 장형사동 구문이 포함된 문장 용례는 모두 450건으로 ‘-게 만들다’ 구문과 ‘-게 하다’(‘-게 해-’, ‘-게 했-’, ‘-게 한’, ‘-게 할’ 등 포함) 구문이 각각 77건과 373건이었다. 전체 장형사동 구문의

용례 450건 가운데 단형사동 형태로 교체할 수 있는 것은 모두 27건으로 나타났다. 이러한 조사 결과는 장형사동 구문의 동사가 단형사동 구문 사동사의 사용 영역을 일정 부분 잠식하고 있으며, 나아가서 일반동사의 사용 영역까지 사용 범위를 확장했음을 알려준다.

원문과의 대조를 통해 장형사동 구문의 연원을 조사한 결과는 다음과 같았다. 원문의 사역동사 구문을 장형사동 구문으로 번역한 용례는 모두 65건(14.5%)이었다. 다음으로 일반 동사를 장형사동 구문으로 번역한 용례가 222건(49.3%)이었다. 끝으로 특정한 동사나 구문의 영향을 받지 않은 상태에서 번역자가 의역의 수단으로 장형사동 구문을 사용한 용례가 163건(36.2%)이었다. 조사 결과를 단순화하면 원문의 사역동사 구문을 번역한 유형(65건, 14.4%)에 비해 사역동사 구문과 무관한 구문을 번역한 유형(385건, 85.6%)이 압도적으로 많았다고 정리할 수 있다.

한국어 번역문에 장형사동 구문이 빈출하는 이유는 원문의 사역동사 구문을 기계적으로 대응 번역한 데서 비롯된 것이 아니라, 사역동사와 일반동사 구문을 번역할 때 원문에 내재된 사역([+causative])이라는 의미기능을 번역문 독자들에게 분명하게 전달하기 위해 한국어에서 동일한 의미기능을 가진 통사 결합(‘-게 만들다’, ‘-게 하다’)으로 처리하는 번역 전략을 채택한 데서 비롯된 것으로 이는 구조적 명시화의 결과로 볼 수 있다.

2.3. 담화 층위 - 문장 접속부사

병렬코퍼스를 활용한 접속부사 ‘그러나’(‘하지만’, ‘그렇지만’)와 ‘그런데’(‘근데’)의 용례 조사를 통해서 확인하고자 한 내용은 다음의 두 가지였다. 첫째는 강한 대조(=역접)를 나타내는 ‘그러나’ 계통 접속부사와 약한 대조(=전환)를 나타내는 ‘그런데’ 계통 접속부사의 상대적 분포를 확인하는 작업으로, 이는 영-한 번역문에서 ‘그런데’와 ‘그러나’의 상대 빈도 차이를 언급한 김혜영(2009a: 64)의 조사 결과(300건 : 43건)를 확인하는 성격을 가졌다. 둘째는 ‘그러나’(‘하지만’ 포함)가 원문의 무엇을 번역한 결과인가를 조사했는데, 이는 우리의 일관된 관심사인 번역 중재 과정을 알아보는 동시에, 영-한 번역문에 대조 접속의 쓰임이 많은 이유를 원문의 대조 접속사 ‘but’을 번역하는 과정에서 전환 관계

의 쓰임을 강한 대조로 잘못 번역한 데서 찾은 김혜영(2009a: 63-68, 133)의 주장을 검증하려는 의도였다.

우리가 구축한 한국어 번역문 코퍼스에서 문장 접속부사가 포함된 문장 용례는 모두 808건으로 ‘그러나’와 ‘하지만’, ‘그렇지만’, ‘그런데’, ‘근데’가 나타난 용례는 각각 315건, 370건, 3건, 118건 및 2건이었다. 이어서 직접적 대조 기능의 접속부사 유형과 간접적 대조 기능의 접속부사 유형에 대해 그 상대 빈도를 비교해보면, 대략 $0.174(118+2=120/315+370+3=688)$ 로 나타났다. 이러한 결과는 기존의 조사 결과와 대체적으로 유사한 경향을 노정하고 있었다.

원문과의 대조를 통해 이들 문장 접속부사의 번역 연원을 조사한 결과는 다음과 같았다.

(가) ‘그러나’의 경우를 보면, 원문의 접속사 ‘but’을 번역한 용례가 가장 많았고(129건, 41.0%), 다음으로 원문에 대응하는 어구가 없는데도 문맥에서 유추하여 번역한 용례가 많았으며(105건, 33.3%), 접속사 ‘however’를 번역한 용례가 뒤를 이었다(54건, 17.1%). 또 접속사 ‘yet’을 번역한 용례(10건, 3.2%)와 ‘and’를 번역한 용례(6건, 2%)가 있었고, 원문에 아예 없는 문장을 운색으로 집어넣으면서 들어간 용례(5건, 1.6%)와 문장부호 줄표를 번역한 용례(3건, 0.9%) 및 기타(3건, 0.9%)도 보였다.

(나) ‘하지만’의 경우를 보면, 원문의 접속사 ‘but’을 번역한 용례가 가장 많았고(239건, 64.6%), 다음으로 문맥에서 유추하여 번역한 용례가 많았으며(77건, 20.8%), 접속사 ‘however’를 번역한 용례가 뒤를 이었다(16건, 4.3%). 그 뒤로 접속사 ‘though’를 번역한 용례(13건, 3.5%), 접속사 ‘and’를 번역한 용례(11건, 3.0%), 원문에 아예 없는 문장을 운색으로 집어넣으면서 들어간 용례(6건, 1.6%) 및 접속사 ‘yet’를 번역한 용례(3건, 0.8%)도 보였다.²⁾

(다) ‘그런데’의 경우를 보면, ‘그러나’와 ‘하지만’의 경우와 조금 달리 문맥에서 유추하여 번역한 용례가 가장 많았고(54건, 45.8%), 원문의 접속사 ‘but’을 번역한 용례가 다음으로 많았으며(33건, 28.0%), 접속사 ‘and’를 번역한 용례가 그 뒤를 이었다(18건, 15.2%).³⁾ 이상의 조사 결과를 요약하면, ‘그러나’와 ‘하지

2) 나머지 5건(1.4%)은 유형화가 곤란한 경우였다.

3) 나머지 용례에 대해서는 김정우(2011: 100)의 각주 40) 참고.

만'은 접속사 'but'을 번역한 용례가 가장 많았고(각각 129건, 41.0%; 239건, 64.6%) '그런데'는 문맥에서 유추해서 도입한 용례가 가장 많았다(54건, 45.8%)고 말할 수 있다.

번역자가 원문에 상응하는 언어적 대응 형태가 없는 상태에서 문맥 지식에 의존해서 '그러나'와 '하지만' 및 '그런데'로 번역한 용례가 상당히 높은 빈도로 출현하고 있다는 사실은 번역문 독자들에게 원문의 의미를 보다 분명하게 전달하기 위해, 대표적 연결 표현인 문장 접속부사를 번역문에 도입한 데서 비롯된 선택적 명시화의 결과로 볼 수 있다.

2.4. 문장부호 - 줄표

병렬코퍼스를 활용하여 문장부호 줄표(dash)의 번역 양상을 조사하는 목적은 영어의 줄표를 한국어로 번역한 기술적 자료의 확보에 있는바, 이는 양쪽 언어에서 동일한 외형을 보이는 문장부호가 실제 번역 작업에서 어떻게 전환되고 있는가를 조사하고, 이를 토대로 영한 번역에서 유용한 줄표의 번역 전략을 수립하고자 하는 의도에서 비롯되었다.

우리가 구축한 영어 원문 코퍼스에서 문장부호 줄표가 포함된 문장 용례는 모두 1,039건이었다. 영어 원문의 줄표는 한국어 번역문에 대략 일곱 가지 유형으로 나타났다. 가장 많은 것은 줄표를 전후로 문장이 분리되어 새로운 문장이 생성되는 용례였다(421건). 다음으로 많은 것이 줄표를 없애고 어순을 재배치하거나(203건) 줄표를 없애고 어순을 그대로 둔 채 풀어쓰는(188건) 용례였다. 기타 줄표가 있는 어구를 번역자가 생략하거나 누락한 용례(116건)⁴⁾, 말줄임표로 변환한 용례(74건), 괄호로 처리한 용례(24건), 쉼표로 대체한 용례(13건) 등이 나타났다.

위의 조사 결과에서 주목되는 것은 원문의 줄표가 동일한 형태의 한국어 문장부호 줄표로 나타나는 용례가 단 하나도 없었다는 사실이다. 이는 원문 특유의 텍스트 특징을 번역문의 전형적 텍스트 특징에 순응하도록 만든 규범화(표준화)의 사례로 볼 수 있다.

4) 해당 줄표가 있는 문장이나 문단을 통째로 누락시킨 용례도 여기에 포함했다.

2.5. 분석 방법과 절차

우리는 앞서 25만 어절 규모의 영-한 병렬코퍼스 분석 결과를 2.1절에서 2.4절에 걸쳐 요약적으로 제시했다. 여기에 요약된 분석 결과를 기준으로 우리는 앞으로 병렬코퍼스의 규모를 15만 어절씩 누적시키면서 이에 따라 분석 결과가 어떻게 변화하는가를 관찰하여, 코퍼스의 규모가 어느 정도까지 되어야 의미 있는 일반화를 담보할 수 있는가를 탐색하게 될 것이다. 논지 전개의 투명성을 확보하기 위해 변화 추이의 관찰과 분석 대상이 되는 항목과 범위를 다시 한 번 간략하게 규정하기로 한다.

(가) 의존명사 ‘때문’

영-한 번역문에서 ‘때문’이 출현하는 환경(분포)을 조사하여 상대 빈도를 확인하고, 원문과의 대조를 통해 번역 연원(영형태, 접속사, 전치사, 특정 구문, 문장부호, 명사나 부사)의 상대 빈도를 산출하여 25만 어절을 조사해서 얻었던 선택적 명시화 경향의 변화 추이를 확인한다.

(나) 장형사동 구문

장형사동 구문의 용례 가운데서 단형사동 구문으로 교체할 수 있는 용례가 얼마나 되는지를 알아보고, 원문과의 대조를 통해 번역 연원(사역동사, 일반동사, 의역)을 조사하고 상대 빈도를 산출하여 25만 어절을 조사해서 얻었던 구조적 명시화 경향의 변화 추이를 확인한다.

(다) 문장 접속부사

직접 대조를 나타내는 접속부사(‘그러나’와 ‘하지만’)⁵⁾와 간접 대조를 나타내는 접속부사(‘그런데’)의 상대적 분포를 확인하고, 원문과의 대조를 통해 문장 접속부사 ‘그러나’와 ‘하지만’, ‘그런데’의 번역 연원을 조사하고 그 상대 빈도를 산출하여 25만 어절을 조사해서 얻었던 선택적 명시화 경향의 변화 추이

5) 전자는 표층에서 드러나는 직접 대조의 용법이고 후자는 표층에서 드러나지 않는 심층 직접 대조의 용법이다. 한편 ‘그렇지만’과 ‘근데’는 표본의 수효가 많지 않아 조사 결과의 정리 과정에서 제외했다.

를 확인한다.

(라) 줄표

영어 원문의 줄표가 한국어 번역문에서 어떻게 재현되었는가를 조사하여, 그 연원(7~8가지 유형)의 상대 빈도를 산출하여 25만 어절을 조사해서 얻었던 규범화 경향의 변화 추이를 확인한다.

3. 코퍼스 크기의 증가에 따른 분석 결과의 변화

3.1. 누적 40만 어절

새로 추가된 15만 어절에 대한 분석을, 25만 어절을 대상으로 한 기존의 분석에 누적시킨 결과는 다음과 같다.

<코퍼스 목록>6)

6. 『모비딕』, 229쪽, 35,000어절, *Moby Dick*, 푸른숲, 2007년.

7. 『세계는 평평하다』, 816쪽, 115,000어절, *The World Is Flat*, 창해, 2006년.

6) 자료의 분류 번호는 25만 어절 조사에 활용했던 다섯 권에 이어서 번호를 매기기로 한다. 자료의 누적에 따른 각종 변수의 변화 양상을 알아보기 위한 조사이므로 누적 일련번호를 사용한 것이다. 한편 이하 예문에서 괄호 안의 숫자는 순서대로 해당 책에 나오는 용례의 일련번호, 책의 분류번호, 해당 용례가 등장하는 책의 쪽수를 각각 가리킨다. 정성적 분석의 일반화를 위한 코퍼스 크기를 살펴보기 위해 문학과 비문학 텍스트가 섞인 균형 코퍼스를 선택하였고, 이를 위해 일정한 크기의 여러 텍스트를 선별하여 코퍼스로 사용하였다. 하나의 텍스트만을 선택한다면 해당 텍스트에만 종속되는 결과만이 나올 수 있기 때문이다. (익명의 심사자 지적) 그러나 뒤의 각주 31)에 언급했듯이 본고의 코퍼스가 완전한 균형이 잡힌 텍스트라고 말하기는 어려운 것이 사실이다.

3.1.1. 의존명사 ‘때문’

새로운 15만 어절 코퍼스에서 추출한 ‘때문’의 용례는 모두 515건(6-46건, 7-469건)이었다. 이들의 분포는 (1) ‘-기 때문-’ 구성이 372건(6-20건, 7-352건), (2) ‘명사/대명사+때문’ 구성이 92건(6-21건, 7-71건), (3) ‘관형사+때문’ 구성이 50건(6-5건, 7-45건), (4) 문두의 단독형 ‘때문’ 구성이 1건(6-0건, 7-1건) 등으로 나타났다. 이 수치를 앞선 25만 어절에 대한 조사 결과와 합산하면, (1) ‘-기 때문-’ 구성이 789건(417+372건), (2) ‘명사/대명사+때문’ 구성이 223건(131+92), (3) ‘관형사+때문’ 구성이 60건(10+50), (4) 문두(文頭)의 단독형 ‘때문’ 구성이 2건(1+1)이 된다.

[표1] ‘때문’의 분포 (25만 어절과 40만 어절 비교)

유형	25만 어절		누적 40만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1)	417회	74.6%	789회	73.5%
(2)	131회	23.4%	223회	20.7%
(3)	10회	1.8%	60회	5.6%
(4)	1회	0.2%	2회	0.2%

위의 표에서 상대 빈도를 보면 25만 어절에서 40만 어절로 늘어나면서 유형 (1) ‘-기 때문-’ 구성과 (2) ‘명사/대명사+때문’ 구성은 약간 감소한 반면, 유형 (3) ‘관형사+때문’ 구성이 증가한 사실을 확인할 수 있다.

새로운 15만 어절 병렬코퍼스에서 원문과의 대조를 통해 ‘때문’의 번역 연원을 조사한 결과는 다음과 같았다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 125건(6-20건, 7-105건), ② 접속사를 번역한 용례가 235건(6-4건, 7-231건), ③ 전치사를 번역한 용례가 50건(6-7건, 7-43건), ④ 특정 구문(관용구)을 번역한 용례가 89건(6-15건, 7-74건), ⑤ 문장부호를 번역한 용례가 7건(6-0건, 7-7건), ⑥ 명사를 번역한 용례가 7건(6-0건, 7-7건), ⑦ 부사를 번역한 용례가 1건(6-0건, 7-1건) 등으로 나타났다. 아래에 용례를 제시한다.

- (1.T) 선착장 아래로 내려가 그 배를 보고 나는 깜짝 놀랐다. 적어도 삼십
 년은 더 돼 보이는 낡은 배였기 때문이다. (6-6-39)
- (1.S) When I got down to the dock, I was amazed at the sight of her.
 She was at least fifty years old.
- (2.T) 9월 11일은 궁극적으로 미국이 아닌 그들, 그 못된 자들의 날이기
 때문이다. (462-7-785)
- (2.S) Because ultimately September 11 is about *them*—the bad guys—not
 about us.
- (3.T) 나는 켄이 사업 때문에 고민이 많다는 걸 눈치챘다. (351-7-605)
- (3.S) I could tell that Ken had a lot on his mind regarding his business.
- (4.T) 그러나 어스레한 빛 때문에 그들의 얼굴을 분간할 수는 없었다.
 (9-6-48)
- (4.S) The light was too dim to see their faces.
- (5.T) 창조력의 핵심은 여기에 있습니다. 사람들이 머리가 좋아서가 아닙니
 다. 환경 때문입니다. 생각의 자유가 있는 환경 말입니다. (373-7-633)
- (5.S) The nucleus of creativity is here, not because people are smarter—it
 is the environment, the freedom of thought.
- (6.T) Y2K 문제는 컴퓨터가 만들어질 때 시계가 내장되기 때문에 생겼다.
 (77-7-189)
- (6.S) The Y2K bug was a result of the fact that when computers were
 built, they came with internal clocks.
- (7.T) “미국이 이제 우리나라 정부보다 더 직간접적으로 내 삶에 영향을
 미치고 있다. 그렇기 때문에 나는 이제 미국 권력에 영향을 미칠 수
 있도록, 이를테면 투표권을 행사하고 싶다.” (392-7-670)
- (7.S) “If America is now touching my life directly or indirectly more
 than my own government, then I want to have a vote in America’s
 power.”

위의 예문 (1.T)-(7.S)는 각각 앞서 언급한 유형 ①부터 ⑦까지에 해당하는
 용례이다. 이 조사 결과를 25만 어절에 대한 조사 결과와 합산해서 표로 만들

7) 필요한 곳에 밑줄을 그어 원문과 번역문에 표시한 관계로 개별 용례에 대한 설명은
 생략한다. (1.S)에서는 원문의 첫째 문장과 둘째 문장 사이에 아무런 접속 표현이 없
 는데도 번역문 (1.T)에 나타난 ‘때문’에 주목하면 된다. 그리고 (5.S)에서는 출표에 밑
 줄이 그어진 것으로 보면 된다.

면 아래와 같다.⁸⁾

[표 2] '때문'의 번역 연원 (25만 어절과 40만 어절 비교)

유형	25만 어절		누적 40만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
① 영형태	249회	44.5%	375회	34.9%
② 접속사	223회	39.9%	458회	42.6%
③ 전치사	40회	7.2%	90회	8.4%
④ 특정 구문	19회	3.4%	108회	10.1%
⑤ 문장부호	17회	3.0%	24회	2.2%
⑥ 명사/부사	11회	2.0%	19회	1.8%

위의 표에서 상대 빈도를 보면 25만 어절에서 40만 어절로 늘어나면서 용례 ①과 ② 및 ③과 ④의 순서가 뒤바뀌고 양쪽에서 비교적 큰 폭의 증감이 일어난 사실을 확인할 수 있다.

3.1.2. 장형사동 구문

새로운 15만 어절 코퍼스에서 추출한 장형사동 구문의 용례는 모두 327건(6-10건, 7-317건)이었다. 이들 용례 중에서 장형사동을 단형사동으로 교체할 수 있는 경우는 15건(4.6%)으로 나타났는데,⁹⁾ 이 수치를 앞선 25만 어절에 대한 조사 결과와 합산하면 모두 42건(54.%, 42/777)이 된다.

8) [표 2]에서 용례 ⑦은 개체 수효가 적어서 ⑥으로 합쳤다. 경향을 파악하는 데는 별 무리가 없을 것이다.

9) 사역의 뜻을 가진 동사 '시키다'로 교체할 수 있는 용례가 27건 조사되었는데, 장형사동 구문의 확산 현상을 엿볼 수 있는 자료이다. (여기서는 '시키다'로 교체할 수 있는 용례는 단형사동으로 교체 가능한 용례에서 제외했다.)

[표3] 단형사동 교체 가능 용례 (25만 어절과 40만 어절 비교)

유형	25만 어절		누적 40만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
장형→ 단형	27회	6%	42건	5.4%

위의 표에서 상대 빈도를 보면 25만 어절에서 40만 어절로 늘어나면서 단형사동으로 교체할 수 있는 장형사동 구문의 상대 빈도가 약간 낮아진 것을 확인할 수 있다.

새로운 15만 어절 코퍼스에서 원문과의 대조를 통해 장형사동 구문의 번역 연원을 조사한 결과는 다음과 같았다. ① 사역동사 ‘have’를 번역한 용례가 2건(6-0건, 7-2건), ② 사역동사 ‘make’를 번역한 용례가 51건(6-1건, 7-50건), ③ 사역동사 ‘let’을 번역한 용례가 11건(6-1건, 7-10건), ④ 사역동사 ‘get’을 번역한 용례가 7건(6-0건, 7-7건), ⑤ 사역의 의미자질을 가진 일반동사를 번역한 용례가 50건(6-0건, 7-50건)¹⁰⁾, ⑥ 일반동사를 번역한 용례가 111건(6-8건, 7-103건), ⑦ 의역에서 비롯된 용례가 95건(6-0건, 7-95건) 등으로 나타났다. 아래에 용례를 제시한다.

- (8.T) (그러므로 전쟁을 벌여 공급사슬에서 떨어지는 것은) 유전을 바닥나게 하거나 폐쇄하는 것과 같은 꼴이다. (121-7-741)
- (8.S) (because you start a war) like having your oil wells go dry or having someone pour cement down them.
- (9.T) 그러나 세계를 좁아지게 만든 그 인프라를 우리도 갖고 있다는 사실을 알게 되었고, 그 사실을 알게 되자 우리는 신속하게 그것을 최대한 활용하기 위해 노력했습니다.(93-7-309)
- (9.S) But we once saw we had an infrastructure which made the world a small place, we promptly tried to make the best use of it.
- (10.T) 이들로 하여금 (방갈로르에서) 간단한 머리기사를 쓰고 표를 만들어 보게 하자. (3-7-34)
- (10.S) Let's let them just do the flash headlines and the tables.

10) 문법적으로 사역동사가 아니지만 사역의 의미를 가진 동사(help, cause, render, enable, allow, force, set, lend 등)를 말한다.

- (11.T) 첫째, 제조업자들과 협력하여 제조원가를 최대한 줄이게 한 것이다.
(55-7-224)
- (11.S) The first was working with the manufacturers to get them to cut their costs as much as possible.
- (12.T) 뿐만 아니라 많은 사람들로 하여금 다른 방식으로 세계를 생각하게 하고 세계를 연속된 전체로 보게 하는 계기가 되었다. (10-7-75)
- (12.S) It is also allowed to think about the world differently—to see it as more of a seamless whole.
- (13.T) “하지만 난 자네 기분을 상하게 하려는 게 아니야.” (7-6-140)
- (13.S) “But I don’t mean to offend, sir.”
- (14.T) 하지만 채권을 언제 사고팔 것인가를 판단하는 것은 자신만의 독특한, 모두를 깜짝 놀라게 할 만한 알고리즘이며 결국 사업의 성패를 결정하게 되는 부분이다. (33-7-128)
- (14.S) It will be your unique, knock-their-socks-off algorithm for when to buy and sell these bonds that will ultimately determine your success or failure.

위의 예문 (8.T)-(14.S)는 각각 앞서 언급한 유형 ①부터 ⑦까지에 해당하는 용례이다. 이 조사 결과를 25만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.¹¹⁾

[표 4] 장형사동 구문의 번역 연원 (25만 어절과 40만 어절 비교)

유형	25만 어절		누적 40만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) 사역동사	65회	14.5%	136회	17.5%
(2) 일반동사	222회	49.3%	383회	49.3%
(3) 의역	163회	36.2%	258회	33.2%

11) [표 4]에서는 조사 유형을 크게 사역동사를 번역한 용례와 일반동사를 번역한 용례 및 의역에서 비롯된 용례 등 셋으로 나누었다. 우리의 조사 목표가 한국어 번역문의 구조적 명시화와 장형사동 구문 용법의 확산 경향을 확인하는 것이므로, 세 가지 유형화만으로도 소기의 성과를 거둘 수 있으리라고 판단했다.

위의 표에서 상대 빈도를 보면 25만 어절에서 40만 어절로 늘어나면서 용례의 순서는 그대로이지만 (1)과 (3)의 비율에 약간의 증감이 있다는 사실을 확인할 수 있다.

3.1.3. 접속부사

새로운 15만 어절 코퍼스에서 추출한 접속부사 ‘그러나’, ‘하지만’, ‘그런데’ 등의 용례는 모두 872건(6-156건, 7-716건)이었다. 이들 용례를 직접 대조(‘그러나’와 ‘하지만’)와 간접 대조(‘그런데’)로 나누어 그 출현 횟수를 비교해보면 808건(598+210)과 64건으로 나타났다. 이 수치를 앞선 25만 어절에 대한 조사 결과와 합산하면, 각각 1,493건(685+808)과 182건(118+64)으로 나타났다. 이 조사 결과를 도표로 만들면 다음과 같다.

(표 5) 직접 대조와 간접 대조의 빈도 (25만 어절과 40만 어절 비교)

유형	25만 어절		누적 40만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
간접/직접	118/685	17.2%	182/1493	12.2%

위의 도표를 보면 직접 대조와 간접 대조 용법의 비율이 25만 어절일 때보다 40만 어절일 때 감소했음을 알 수 있다.

새로운 15만 어절 병렬코퍼스에서 원문과의 대조를 통해 ‘그러나’와 ‘하지만’ 및 ‘그런데’의 번역 연원을 조사한 결과는 다음과 같았다.

3.1.3.1. ‘그러나’

먼저 598건의 용례가 조사된 ‘그러나’의 경우를 살펴본다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 91건(6-13건, 7-78건), ② 접속사 ‘but’를 번역한 용례가 436건(6-71건, 7-365건), ③ 접속사 ‘however’를 번역한 용례가 11건(6-0건, 7-11건), ④ 접속사 ‘yet’를 번역한 용례가 7건(6-0건, 7-7건), ⑤ 접속사 ‘and’를 번역한 용례가 19건(6-3건, 7-16건), ⑥ 접속사 ‘though’를 번역한 용례가 24건(6-0건, 7-24건) ⑦

문장부호를 번역한 용례가 7건(6-0건, 7-7건), ⑧ 기타의 용례가 3건(6-3, 7-0)¹²⁾ 등으로 나타났다. 아래에 용례를 제시한다.

- (15.T) 그러나 여유만만하고 능숙한 그의 태도에는 뭐랄까, 왕과 같은 위엄이 스며 있었다. (10-6-33)
- (15.S) He was so relaxed and masterful doing this, there was something almost kingly about it.
- (16.T) 굳이 만나야 할 필요도 없다. 그러나 이진 나름대로 괜찮은 일자리이고 수입도 그만하면 적당하고 법인에서도 내게 별 불만이 없는 눈치다. (9-7-30)
- (16.S) They keep me in the back. But it is a decent living and the firm is basically happy with me.
- (17.T) (위키피디아는 이런 방식으로 작동한다.) 그러나 기술 그 자체만으로는 충분하지 않다. (82-7-175)
- (17.S) (Wikipedia works by ~.) However, the technology is not enough on its own.
- (18.T) 오늘날의 사우디아라비아는 단지 석유 판매로만 먹고 산다. 그러나 사우디아라비아 바로 옆에는 오일 달러로 걸프 만에 무역, 관광, 서비스, 컴퓨터 센터를 세운 도시국가 두바이가 있다. (343-7-590)
- (18.S) Saudi Arabia today can get by just selling oil. Yet right next to Saudi Arabia sits Dubai, an Arab city-state that has used its petrodollars to build the trading, tourist, service, and computing center of the Arab Gulf.
- (19.T) “나는 이 지역에서는 크다고 할 수 있지만 세계적인 규모로 보면 작았습니다. 그러나 나는 그것을 뒤집었지요.” (368-7-619)
- (19.S) “I was big locally and small internationally – and I reversed it.”
- (20.T) 그러나 위키피디아에 언제나 눈부시게 빛나는 달콤한 현실만 존재하는 것은 물론 아니다. (84-7-177)
- (20.S) Wikipedia, though, is not all sweetness and light.
- (21.T) 물론 돈을 많이 벌기 위해 그럴 수도 있고요. 그러나 단지 지식세계에 공헌하려는 목적인 경우도 있습니다. (69-7-138)
- (21.S) Sometimes you can build a business out of it; sometimes they just want to increase the store of knowledge in the world.

12) 원문과 사뭇 다르게 내용을 윤색하면서 ‘그러나’가 들어간 경우 등이다.

위의 예문 (15.T)-(21.S)는 각각 앞서 언급한 유형 ①부터 ⑦까지에 해당하는 용례이다. 이 조사 결과를 25만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.¹³⁾

[표 6] '그러나'의 번역 연원 (25만 어절과 40만 어절 비교)

유형	25만 어절		누적 40만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) 'but'	129회	41.0%	565회	61.9%
(2) 영형태	105회	33.3%	196회	21.5%
(3) 'however'	54회	17.1%	65회	7.1%

위의 표에서 상대 빈도를 보면 25만 어절에서 40만 어절로 늘어나면서 유형 (1)의 상대 빈도는 큰 폭으로 증가한 반면, 유형 (2)와 (3)의 상대 빈도는 대략 유형 (1)에서 증가한 분량의 절반만쯤 감소했음을 확인할 수 있다.

3.1.3.2. '하지만'

다음으로 210건의 용례가 조사된 '하지만'의 경우를 살펴본다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 51건(6-5건, 7-46건), ② 접속사 'but'를 번역한 용례가 139건(6-33건, 7-106건), ③ 접속사 'however'를 번역한 용례가 0건, ④ 접속사 'yet'를 번역한 용례가 1건(6-0건, 7-1건), ⑤ 접속사 'and'를 번역한 용례가 8건(6-0건, 7-8건), ⑥ 접속사 'though'를 번역한 용례가 7건(6-0건, 7-7건) ⑦ 문장부호를 번역한 용례가 0건, ⑧ 기타의 용례가 4건(6-4, 7-0)등으로 나타났다. 아래에 용례를 제시한다.

13) 일곱 가지 유형 전부가 아니라, 표본 수효가 비교적 큰 상위 3개만을 대상으로 조사했다. 따라서 도표의 백분율이 100%에 못 미친다. 이하 '하지만'과 '그러나'의 경우 및 누적 55만, 70만, 85만, 100만 어절의 경우도 동일하다.

- (22.T) 열여섯 명의 사내들이 정신없이 노를 저었다. 하지만 그 거대한 사체를 피쿼드호 옆으로 끌어오는 데에는 대여섯 시간이나 걸렸다. (18-6-96)
- (22.S) Even with sixteen men pulling on the oars, it took us several hours to drag the massive corpse to the side of the sea.
- (23.T) 그의 행동을 지켜보며 나는 더 이상 놀랄 일이 없을 거라고 생각했다. 하지만 그건 오산이었다. (3-6-32)
- (23.S) I didn't think there was anything left to surprise me after watching this performance, but I was mistaken.
- (24.T) 하지만 경기가 끝날 무렵 그 남자가 들고 나가는 돈다발을 보면 경기장 구내에 있는 어떤 상점보다 매상이 좋아 보인다. (98-7-413)
- (24.S) Yet I always notice that by the end of the game he is carrying around a was of bills—and tips—that is thicker than that of any other vendor I see.
- (25.T) (~) 새빨간 거짓말이 이슬람 세계 어딘가에서 시작되어 인터넷에서 들불처럼 번져나갔다. 하지만 그 실체를 밝혀 소문을 불식시킬 수 있는 뉴스는 어디에도 없었다. (81-7-332)
- (25.S) The big lie (~) began somewhere in the Muslim world and spread like a wildfire on the Internet, and no amount of news stories debunking it could eradicate that rumor.
- (26.T) 하지만 기자의 신분이니만큼 나의 이 질문에 정면으로 맞서 해결책을 찾고자 애써온 진정한 교육자가 무수히 많다는 사실을 전달할 수는 없다. (110-7-437)
- (26.S) I am a reporter, though, and I can report that there are some real educators out there who have tried to address this question head-on.

위의 예문 (22.T)-(26.S)는 각각 앞서 언급한 유형 ①, ②, ④, ⑤, ⑥에 해당하는 용례이다. 이 조사 결과를 25만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

(표 7) '하지만'의 번역 연원 (25만 어절과 40만 어절 비교)

유형	25만 어절		누적 40만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) 'but'	239회	64.6%	378회	65.2%
(2) 영형태	77회	20.8%	128회	22.1%
(3) 'though'	13회	3.5%	20회	3.4%
(4) 'and'	11회	3.0%	19회	3.2%

위의 표에서 상대 빈도를 보면 25만 어절에서 40만 어절로 늘어났지만 각 용례의 상대 빈도는 1~2% 차이 수준에서 거의 비슷하게 유지되고 있음을 확인할 수 있다.

3.1.3.3. '그런데'

마지막으로 64건의 용례가 조사된 '하지만'의 경우를 살펴본다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 23건(6-7건, 7-16건), ② 접속사 'but'를 번역한 용례가 22건(6-12건, 7-10건), ③ 접속사 'however'를 번역한 용례가 0건, ④ 접속사 'yet'를 번역한 용례가 1건(6-0건, 7-1건), ⑤ 접속사 'and'를 번역한 용례가 15건(6-4건, 7-11건), ⑥ 접속사 'though'를 번역한 용례가 1건(6-0건, 7-1건) ⑦ 문장부호를 번역한 용례가 0건, ⑧ 기타의 용례가 2건(6-1, 7-1) 등으로 나타났다. 아래에 용례를 제시한다.

- (27.T) 파이프 담배를 피우려면 무엇보다 마음이 편안하고 행복해야 돼. 일과를 마친 뒤, 조금 뒤에 있을 밤의 휴식을 위해 난롯불 옆 안락 의자에 앉아서라면 금상첨화겠지. 그런데 그런 평온은 나 같은 사람에게겐 아무래도 무리야. (6-6-56)
- (27.S) Pipes are for the happy, contented man for the comfy chair by the fire, for the day's work over and a night's rest to come. That serenity is not for the likes of me.
- (28.T) 미국 상무부에 따르면 미국인이 투자한 해외기업 생산물의 90퍼센

트가 해외시장에서 소화된다. 그런데 이는 미국의 수출을 늘려주기도 한다. (10-7-208)

- (28.S) According to the U.S. Commerce Department, nearly 90 percent of the output from U.S.-owned offshore factories is sold to foreign countries. But this actually stimulates American exports.
- (29.T) 주식가격은 정점에서 떨어진 지 오래고 시장은 침체 국면에 접어들었습니다. 그런데 기업의 전산망 백업 작업에만 많은 비용이 들었죠. (9-7-195)
- (29.S) Everyone found prices flat to declining and markets stagnant, yet they were still spending huge amounts of money running the backroom operations of their companies.
- (30.T) 나는 그의 팔을 치우려고 몸을 뒤틀어 보았다. 그런데 무언가가 다리 근처에 걸리적이었다. (2-6-31)
- (30.S) I twisted away from him and felt a scratch on my leg.
- (31.T) 그런데 이 얘기를 들으면 어떤 느낌이 들까? 롤스로이스는 이제 더 이상 자동차를 만들지 않는다는 것 말이다. (34-7-625)
- (31.S) What if I told you, though, Rolls-Royce even doesn't make cars any more.

위의 예문 (27.T)-(31.S)는 각각 앞서 언급한 유형 ①, ②, ④, ⑤, ⑥에 해당하는 용례이다. 이 조사 결과를 25만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

[표 8] '그런데'의 번역 연원 (25만 어절과 40만 어절 비교)

유형	25만 어절		누적 40만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
① 영형태	54회	45.8%	77회	42.3%
② 'but'	33회	28.0%	55회	30.2%
③ 'and'	18회	15.2%	33회	18.1%

위의 표에서 상대 빈도를 보면 25만 어절에서 40만 어절로 늘어나면서 ①에서는 약간 감소한 반면, ②와 ③에서는 약간 증가했음을 확인할 수 있다.

3.1.4. 줄표

우리가 새로 구축한 15만 어절의 영어 원문 코퍼스에서 문장부호 줄표가 포함된 문장 용례는 모두 1,086건이었다. 영어 원문의 줄표는 한국어 번역문에 대략 일곱 가지 유형으로 나타났다. ① 줄표를 전후로 문장이 분리되어 새로운 문장이 생성되는 용례가 410건(6-18, 7-392), ② 줄표를 없애고 어순을 재배치한 용례가 227건(6-8, 7-219), ③ 줄표를 없애고 어순을 그대로 둔 채 풀어쓴 용례가 261건(6-13, 7-248), ④ 줄표가 있는 어구를 번역자가 누락한 용례가 68건(6-5, 7-63), ⑤ 말줄임표로 변환한 용례가 1건(6-0, 7-1), ⑥ 괄호로 처리한 용례가 119건(6-0, 7-119), ⑦ 쉼표로 대체한 용례가 0건, ⑧ 줄표로 변환한 용례가 0건 등으로 나타났다. 아래에 용례를 제시한다.¹⁴⁾

- (32.S) At first I thought he'd been in a fight and was terribly bruised and cut—the landlord had put me in bed with a back-street brawler. (8-6-21)
- (32.T) 처음에는 그저 싸우다가 다쳐 생긴 상처이겠거니, 하고 여겼다. 여관 주인이 나를 뒷골목의 강패와 같은 방에 넣었다는 생각에 잠시 화가 치밀기도 했다.
- (33.S) Lumber and tools were stashed away, along with every conceivable spare part for the whaleboats themselves—the slender craft that are sent out to chase the whale. (15-6-36)
- (33.T) 목재와 각종 연장, 그리고 앞뒤가 뽕죽하고 몸체가 가느다란 고래잡이용 보트에 꼭 필요하다고 생각되는 부품들은 모두 한 편에 따로 놓아두었다.
- (34.S) It was the “line”—the thick rope that is fixed to the harpoon—taking up the strain. (24-6-69)
- (34.T) 그것은 ‘줄’, 그러니까 작살에 달린 굵은 로프가 힘을 받는 소리였다.
- (35.S) The room was unheated—Coffin said he couldn't afford such luxuries—and as cold as Iceland. (5-6-18)
- (35.T) 방 안에는 온기라곤 전혀 없었다. (ø) 마치 아이슬란드처럼 냉골이었다.

14) 밑줄 그은 부분이 영어 원문의 줄표가 이끄는 어구와 상응하는 곳이다.

- (36.S) Microsoft Windows, Office, 3D Studio MAX, Adobe Photoshop—each of these programs would have been priced higher than they are today (~). (419-7-220)
- (36.T) 마이크로소프트 윈도우, 오피스, 3D 스튜디오, 어도비 포토샵……. 이들 프로그램은 (~) 지금보다 비싼 가격이 매겨졌을 겁니다.
- (37.S) (~ when) the walls started falling all around the world, and integration—and the backlash to it—went to a whole new level. (20-7-10)
- (37.T) 세계 곳곳에서 장벽들이 무너져 내리고 세계화의 과정이 (그리고 그에 따른 반발까지도) 완전히 새로운 단계로 진전된 시기였다.

위의 예문 (32.T)-(37.S)는 각각 앞서 언급한 유형 ①, ②, ③, ④, ⑤, ⑥에 각각 해당하는 용례이다. 이 조사 결과를 25만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

[표 9] 줄표의 번역 양상 (25만 어절과 40만 어절 비교)

유형	25만 어절		누적 40만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
① 새 문장	421회	40.5%	831회	39.1%
② 어순 도치	203회	19.5%	430회	20.2%
③ 어순 유지	188회	18.1%	449회	21.1%
④ 생략/누락	116회	11.2%	184회	8.7%
⑤ 말줄임표	74회	7.1%	75회	3.5%
⑥ 괄호	24회	2.3%	143회	6.8%
⑦ 쉼표	13회	1.3%	13회	0.6%
⑧ 줄표	0회	0.0%	0회	0.0%

위의 표를 보면, ②의 ③ 및 ⑤와 ⑥의 순위가 뒤바뀌었고, ③과 ④, ⑤, ⑥ 등에서 상대 빈도가 약간씩 변화를 보이고 있음을 확인할 수 있다. (여기서도 앞서 25만 어절의 경우와 마찬가지로 줄표를 그대로 가져온 번역 용례는 발견되지 않았다.)

3.2. 누적 55만 어절

2차로 추가된 15만 어절에 대한 분석을, 기존의 누적 40만 어절 분석에 누적시킨 결과를 항목별로 정리하면 다음과 같다.¹⁵⁾

<코퍼스 목록>¹⁶⁾

8. 『렉서스와 올리브나무』, 822쪽, 150,000어절, *Lexus and Olive Tree*, 창해, 2007년.

3.2.1. 의존명사 ‘때문’

새로운 2차 15만 어절 코퍼스에서 추출한 ‘때문’의 용례는 모두 559건이었다. 이들의 분포는 (1) ‘-기 때문-’ 구성이 401건, (2) ‘명사/대명사+때문’ 구성이 65건, (3) ‘관형사+때문’ 구성이 70건, (4) 문두의 단독형 ‘때문’ 구성이 15건 등으로 나타났다.¹⁷⁾ 이 수치를 앞의 누적 40만 어절에 대한 조사 결과와 합산하면, (1) ‘-기 때문-’ 구성이 1,190건(789+401건), (2) ‘명사/대명사+때문’ 구성이 288건(223+65), (3) ‘관형사+때문’ 구성이 130건(60+70), (4) 문두(文頭)의 단독형 ‘때문’ 구성이 17건(2+15)이 된다.

[표10] ‘때문’의 분포 (40만 어절과 55만 어절 비교)

유형	누적 40만 어절		누적 55만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1)	789회	73.5%	1190회	72.9%
(2)	223회	20.7%	288회	17.6%
(3)	60회	5.6%	130회	8.0%
(4)	2회	0.2%	17회	1.0%

15) 55만 어절 이후의 조사 결과 기술에서는 용례를 제시하지 않고, 통계 수치만 제시하기로 한다.

16) 원래 이 책의 코퍼스 크기는 168,000어절이었는데, 양적 변수의 크기를 일정하게 맞추기 위해 10% 정도를 감축해서 150,000어절까지만 조사했다.

17) 위의 4개 유형에 속하지 않는 새로운 유형(용언의 관형형+때문)이 8건 조사되었다. (예) 이는 해당 국가의 내부 개혁에 너무 깊이 관여할 경우, 그 나라가 상대편 쪽으로 돌아서버릴 것을 우려한 때문이었다(329-9-434). 누적 55만 어절을 기준으로 0.5%에 해당하는 분포이다. 그래서 [표 10]의 전체 합계가 100%에 조금 못 미친다.

위의 표에서 상대 빈도를 보면 40만 어절에서 55만 어절로 늘어나면서 (2) ‘명사/대명사+때문’ 구성 유형은 약간 감소한 반면, 그 비율만큼 (3) ‘관형사+때문’ 구성 유형이 증가했음을 확인할 수 있다.

새로운 2차 15만 어절 병렬코퍼스에서 원문과의 대조를 통해 ‘때문’의 번역 연원을 조사한 결과는 다음과 같았다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 161건, ② 접속사를 번역한 용례가 256건, ③ 전치사를 번역한 용례가 30건, ④ 특정 구문을 번역한 용례가 93건, ⑤ 문장부호를 번역한 용례가 10건, ⑥ 명사를 번역한 용례가 9건, ⑦ 부사를 번역한 용례가 0건 등으로 나타났다. 이 조사 결과를 누적 40만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

[표 11] ‘때문’의 번역 연원 (40만 어절과 55만 어절 비교)

유형	누적 40만 어절		누적 55만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
① 영형태	375회	34.9%	536회	32.8%
② 접속사	458회	42.6%	714회	43.7%
③ 전치사	90회	8.4%	120회	7.4%
④ 특정 구문	108회	10.1%	201회	12.3%
⑤ 문장부호	24회	2.2%	34회	2.1%
⑥ 명사/부사	19회	1.8%	28회	1.7%

위의 표에서 상대 빈도를 보면 40만 어절에서 55만 어절로 늘어나면서 유형 ①과 ③이 감소한 반면, ②와 ④가 증가한 사실을 확인할 수 있다.

3.2.2. 장형사동 구문

새로운 2차 15만 어절 코퍼스에서 추출한 장형사동 구문의 용례는 모두 164건이었다. 이들 용례 중에서 장형사동을 단형사동으로 교체할 수 있는 경우는 13건(7.9%)으로 나타났는데, 이 수치를 앞의 누적 40만 어절에 대한 조사 결과와 합산하면 모두 55건(5.8%, 55/941)이 된다. 이는 누적 40만 어절을 대상으로 한 조사에서 얻었던 결과와 엇비슷한 수치이다

[표 12] 단형사동 교체 가능 용례 (40만 어절과 55만 어절 비교)

유형	누적 40만 어절		누적 55만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
장형→ 단형	42회	5.4%	55건	5.8%

새로운 2차 15만 어절 코퍼스에서 원문과의 대조를 통해 장형사동 구문의 번역 연원을 조사한 결과는 다음과 같았다. ① 사역동사 ‘have’를 번역한 용례가 2건, ② 사역동사 ‘make’를 번역한 용례가 19건, ③ 사역동사 ‘let’을 번역한 용례가 6건, ④ 사역동사 ‘get’을 번역한 용례가 2건, ⑤ 사역의 의미자질을 가진 일반동사를 번역한 용례가 0건, ⑥ 일반동사를 번역한 용례가 67건, ⑦ 의역에서 비롯된 용례가 68건 등으로 나타났다. 이 조사 결과를 누적 40만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

[표 13] 장형사동 구문의 번역 연원 (40만 어절과 55만 어절 비교)

유형	누적 40만 어절		누적 55만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) 사역동사	136회	17.5%	165회	17.5%
(2) 일반동사	383회	49.3%	450회	47.8%
(3) 의역	258회	33.2%	326회	34.7%

위의 표에서 상대 빈도를 보면 40만 어절에서 55만 어절로 늘어나면서 유형의 순서는 그대로이지만 유형 (2)가 감소한 반면, 그만큼 유형 (3)이 증가한 사실을 확인할 수 있다.

3.2.3. 접속부사

새로운 2차 15만 어절 코퍼스에서 추출한 접속부사 ‘그러나’, ‘하지만’, ‘그런데’ 등의 용례는 모두 969건이었다. 이들 용례를 직접 대조(‘그러나’와 ‘하지만’)와 간접 대조(‘그런데’)로 나누어 그 출현 횟수를 비교해보면 796건 (419+377)과 173건으로 나타났다. 이 수치를 앞의 누적 40만 어절에 대한 조사

결과와 합산하면, 각각 2,289건(1493+796)과 355건(182+173)으로 나타났다.

[표 14] 직접 대조와 간접 대조의 빈도 (40만 어절과 55만 어절 비교)

유형	누적 40만 어절		누적 55만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
간접/직접	182/1493	12.2%	355/2289	15.5%

위의 도표를 보면 40만 어절에서 55만 어절로 늘어나면서 직접 대조와 간접 대조 용법의 비율이 소폭 증가했음을 알 수 있다.

새로운 2차 15만 어절 병렬코퍼스에서 원문과의 대조를 통해 ‘그러나’와 ‘하지만’ 및 ‘그런데’의 번역 연원을 조사한 결과는 다음과 같았다.

3.2.3.1. ‘그러나’

먼저 모두 419건의 용례가 조사된 ‘그러나’의 경우를 살펴본다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 146건, ② 접속사 ‘but’를 번역한 용례가 240건, ③ 접속사 ‘however’를 번역한 용례가 1건, ④ 접속사 ‘yet’를 번역한 용례가 3건, ⑤ 접속사 ‘and’를 번역한 용례가 10건, ⑥ 접속사 ‘though’를 번역한 용례가 17건, ⑦ 문장부호를 번역한 용례가 2건 등으로 나타났다. 이 조사 결과를 40만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

[표 15] ‘그러나’의 번역 연원 (40만 어절과 55만 어절 비교)

유형	누적 40만 어절		누적 55만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) ‘but’	565회	61.9%	805회	60.4%
(2) 영형태	196회	21.5%	342회	25.7%
(3) ‘however’	65회	7.1%	66회	4.9%

위의 표에서 상대 빈도를 보면 40만 어절에서 55만 어절로 늘어나면서 유

형 (2)의 상대 빈도가 약간 증가하고 (1)과 (3)의 상대 빈도가 감소했음을 확인할 수 있다.

3.2.3.2. ‘하지만’

다음으로 377건의 용례가 조사된 ‘하지만’의 경우를 살펴본다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 131건, ② 접속사 ‘but’를 번역한 용례가 2169건, ③ 접속사 ‘however’를 번역한 용례가 0건, ④ 접속사 ‘yet’를 번역한 용례가 0건, ⑤ 접속사 ‘and’를 번역한 용례가 21건, ⑥ 접속사 ‘though’를 번역한 용례가 4건 ⑦ 문장부호를 번역한 용례가 5건 등으로 나타났다. 이 조사 결과를 누적 40만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

[표 16] ‘하지만’의 번역 연원 (40만 어절과 55만 어절 비교)

유형	누적 40만 어절		누적 55만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) ‘but’	378회	65.2%	594회	62.1%
(2) 영형태	128회	22.1%	259회	27.1%
(3) ‘though’	20회	3.4%	24회	2.5%
(4) ‘and’	19회	3.2%	40회	4.2%

위의 표에서 상대 빈도를 보면 40만 어절에서 55만 어절로 늘어나면서 유형 (1)과 (3)이 줄어든 반면, 유형 (2)와 (4)가 늘어난 사실을 확인할 수 있다.

3.2.3.3. ‘그런데’

마지막으로 173건의 용례가 조사된 ‘하지만’의 경우를 살펴본다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 110건, ② 접속사 ‘but’를 번역한 용례가 29건, ③ 접속사 ‘however’를 번역한 용례가 0건, ④ 접속사 ‘yet’를 번역한 용례가 1건, ⑤ 접속사 ‘and’를 번역한 용례가 27건, ⑥ 접속사 ‘though’를 번역한 용례가 3건 ⑦ 문장부호를

번역한 용례가 3건 등으로 나타났다. 이 조사 결과를 누적 40만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

[표 17] '그런데'의 번역 연원 (40만 어절과 55만 어절 비교)

유형	누적 40만 어절		누적 55만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) 영형태	77회	42.3%	187회	52.7%
(2) 'but'	55회	30.2%	84회	23.7%
(3) 'and'	33회	18.1%	60회	16.9%

위의 표에서 상대 빈도를 보면 40만 어절에서 55만 어절로 늘어나면서 유형 (1)의 비율이 대폭 증가한 반면, 유형 (2)와 (3)의 비율이 그만큼 감소했음을 확인할 수 있다.

3.2.4. 줄표

우리가 새로 구축한 2차 15만 어절의 영어 원문 코퍼스에서 문장부호 줄표가 포함된 문장 용례는 모두 758건이었다. 영어 원문의 줄표는 한국어 번역문에 대략 일곱 가지 유형으로 나타났다. ① 줄표를 전후로 문장이 분리되어 새로운 문장이 생성되는 용례가 413건, ② 줄표를 없애고 어순을 재배치한 용례가 123건, ③ 줄표를 없애고 어순을 그대로 둔 채 풀어쓴 용례가 137건, ④ 줄표가 있는 어구를 번역자가 누락한 용례가 67건¹⁸⁾, ⑤ 말줄임표로 변환한 용례가 0건, ⑥ 괄호로 처리한 용례가 8건, ⑦ 쉼표로 대체한 용례가 5건, ⑧ 줄표로 변환한 용례가 5건 등으로 나타났다. 이 조사 결과를 누적 40만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

18) 이 가운데는 줄표가 포함된 문단 자체가 누락된 용례가 14건, 본문에서 빠진 대신 각주로 처리된 용례가 1건 조사되었다.

[표 18] 줄표의 번역 양상 (40만 어절과 55만 어절 비교)

유형	누적 40만 어절		누적 55만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
① 새 문장	831회	39.1%	1244회	43.2%
② 어순 도치	430회	20.2%	553회	19.2%
③ 어순 유지	449회	21.1%	586회	20.3%
④ 생략/누락	184회	8.7%	251회	8.7%
⑤ 말줄임표	75회	3.5%	75회	2.6%
⑥ 괄호	143회	6.8%	151회	5.2%
⑦ 쉼표	13회	0.6%	18회	0.6%
⑧ 줄표	0회	0.0%	5회	0.2%

위의 표를 보면, 40만 어절에서 55만 어절로 늘어나면서, 유형 ②의 ③의 순위가 뒤바뀌었고, 유형 ①은 상대 빈도가 증가한 반면, ⑤와 ⑥은 감소했으며, 특히 원문의 줄표를 그대로 번역문에 가져온 유형 ⑧이 비율은 미미하지만 새롭게 등장했다는 사실을 확인할 수 있다.

3.3. 누적 70만 어절

3차로 추가된 15만 어절에 대한 분석을, 기존의 55만 어절 분석에 누적시킨 결과를 항목별로 정리하면 다음과 같다.

<코퍼스 목록>

9. 『영혼의 해부』, 357쪽, 67,000어절, *Anatomy of the Spirit*, 한문화, 2003년.
10. 『블랙 선데이』, 566쪽, 83,000어절, *Black Sunday*, 창해, 2006년.

3.3.1. 의존명사 ‘때문’

새로운 3차 15만 어절 코퍼스에서 추출한 ‘때문’의 용례는 모두 382건이었다. 이들의 분포는 (1) ‘-기 때문-’ 구성이 145건(9-87건, 10-58건), (2) ‘명사/대명사+때문’ 구성이 102건(9-86건, 10-16건), (3) ‘관형사+때문’ 구성이 126건

(9-83건, 10-43건), (4) 문두의 단독형 ‘때문’ 구성이 8건(9-4건, 10-4건) 등으로 나타났다.¹⁹⁾ 다시 이 수치를 앞선 55만 어절에 대한 조사 결과와 합산하면, (1) ‘-기 때문-’ 구성이 1,335건(1190+145건), (2) ‘명사/대명사+때문’ 구성이 390건(288+102), (3) ‘관형사+때문’ 구성이 256건(130+126), (4) 문두(文頭)의 단독형 ‘때문’ 구성이 25건(17+8)이 된다.

[표 19] ‘때문’의 분포 (55만 어절과 70만 어절 비교)

유형	누적 55만 어절		누적 70만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1)	1190회	72.9%	1335회	66.3%
(2)	288회	17.6%	390회	19.4%
(3)	130회	8.0%	256회	12.7%
(4)	17회	1.0%	25회	1.2%

위의 표에서 상대 빈도를 보면 55만 어절에서 70만 어절로 늘어나면서 (1) ‘-기 때문-’ 구성이 감소한 반면, (2) ‘명사/대명사+때문-’ 구성과 ‘관형사 + 때문-’ 구성이 증가했음을 확인할 수 있다.

새로운 3차 15만 어절 병렬코퍼스에서 원문과의 대조를 통해 ‘때문’의 번역 연원을 조사한 결과는 다음과 같았다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 84건(9-44, 10-40), ② 접속사를 번역한 용례가 175건(9-128, 10-47), ③ 전치사를 번역한 용례가 73건(9-54, 10-19), ④ 특정 구문을 번역한 용례가 38건(9-23, 10-15), ⑤ 문장부호를 번역한 용례가 7건(9-7, 10-0), ⑥ 명사를 번역한 용례가 4건(9-4, 10-0), ⑦ 부사를 번역한 용례가 1건(9-0, 10-1) 등으로 나타났다. 이 조사 결과를 누적 55만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

19) 앞의 각주 17)에서와 같이 ‘용언의 관형형+때문’ 유형이 1건 조사되었다. (예) 그 자를 살려둔 때문이란 말이오(59-11-259). 지금까지 누적 70만 어절을 기준으로 0.39(반올림 0.4%)에 해당하는 분포이다.

[표 20] '때문'의 번역 연원 (55만 어절과 70만 어절 비교)

유형	누적 55만 어절		누적 70만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
① 영형태	536회	32.8%	620회	30.8%
② 접속사	714회	43.7%	889회	44.1%
③ 전치사	120회	7.4%	193회	9.6%
④ 특정 구문	201회	12.3%	239회	11.9%
⑤ 문장부호	34회	2.1%	41회	2.0%
⑥ 명사/부사	28회	1.7%	32회	1.6%

위의 표에서 상대 빈도를 보면 55만 어절에서 70만 어절로 늘어나면서 대략 유형 ①이 감소한 반면, 유형 ③이 증가했음을 확인할 수 있다.

3.3.2. 장형사동 구문

새로운 3차 15만 어절 코퍼스에서 추출한 장형사동 구문의 용례는 모두 259건이었다. 이들 용례 중에서 장형사동을 단형사동으로 교체할 수 있는 경우는 20건(7.7%)으로 나타났는데, 이 수치를 앞의 누적 55만 어절에 대한 조사 결과와 합산하면 모두 75건(6.3%, 75/1200)이 된다. 이는 누적 55만 어절을 대상으로 한 조사에서 얻었던 결과보다 조금 증가한 수치이다

[표 21] 단형사동 교체 가능 용례 (55만 어절과 70만 어절 비교)

유형	누적 55만 어절		누적 70만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
장형→ 단형	55회	5.8%	75건	6.3%

새로운 3차 15만 어절 코퍼스에서 원문과의 대조를 통해 장형사동 구문의 연원을 조사한 결과는 다음과 같았다. ① 사역동사 'have'를 번역한 용례가 2건(9-1, 10-1), ② 사역동사 'make'를 번역한 용례가 28건(9-15, 10-13), ③ 사역동사 'let'을 번역한 용례가 8건(9-6, 10-2), ④ 사역동사 'get'을 번역한 용례가 2건(9-2, 10-0), ⑤ 사역의 의미자질을 가진 일반동사를 번역한 용례가 26건

(9-22, 10-4), ⑥ 일반동사를 번역한 용례가 119건(9-77, 10-42), ⑦ 의역에서 비롯된 용례가 74건(9-39, 10-35) 등으로 나타났다. 이 조사 결과를 누적 55만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

[표 22] 장형사동 구문의 번역 연원 (55만 어절과 70만 어절 비교)

유형	누적 55만 어절		누적 70만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) 사역동사	165회	17.5%	205회	17.1%
(2) 일반동사	450회	47.8%	595회	49.6%
(3) 의역	326회	34.7%	400회	33.3%

위의 표에서 상대 빈도를 보면 55만 어절에서 70만 어절로 늘어나면서 유형의 순서는 그대로이면서, 소폭이지만 유형 (2)의 비율이 증가한 반면, 유형 (1)과 (3)의 비율이 감소했음을 확인할 수 있다.

3.3.3. 접속부사

새로운 3차 15만 어절 코퍼스에서 추출한 접속부사 ‘그러나’, ‘하지만’, ‘그런데’ 등의 용례는 모두 487건이었다. 이들 용례를 직접 대조(‘그러나’와 ‘하지만’)와 간접 대조(‘그런데’)로 나누어 그 출현 횟수를 비교해보면 443건(322+121)과 72건으로 나타났다. 이 수치를 앞의 누적 55만 어절에 대한 조사 결과와 합산하면, 각각 2,732건(2289+443)과 427건(355+72)으로 나타났다.

[표 23] 직접 대조와 간접 대조의 빈도 (55만 어절과 70만 어절 비교)

유형	누적 55만 어절		누적 70만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
간접/직접	355/2289	15.5%	427/2732	15.6%

위의 도표를 보면 표본의 크기가 55만 어절에서 70만 어절로 늘어났지만 직접 대조와 간접 대조 용법의 비율은 거의 변화를 보이지 않았음을 알 수 있다.

새로운 3차 15만 어절 병렬코퍼스에서 원문과의 대조를 통해 ‘그러나’와 ‘하지만’ 및 ‘그런데’의 번역 연원을 조사한 결과는 다음과 같았다.

3.3.3.1. ‘그러나’

먼저 모두 322건의 용례가 조사된 ‘그러나’의 경우를 살펴본다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 77건(9-13, 10-64), ② 접속사 ‘but’를 번역한 용례가 162건(9-71, 10-91), ③ 접속사 ‘however’를 번역한 용례가 46건(9-43, 10-3), ④ 접속사 ‘yet’를 번역한 용례가 18건(9-18, 10-0), ⑤ 접속사 ‘and’를 번역한 용례가 11건(9-2, 10-9), ⑥ 접속사 ‘though’를 번역한 용례가 6건(9-4, 10-2), ⑦ 문장부호를 번역한 용례가 2건(9-0, 10-2) 등으로 나타났다. 이 조사 결과를 55만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

[표 24] ‘그러나’의 번역 연원 (55만 어절과 70만 어절 비교)

유형	누적 55만 어절		누적 70만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) ‘but’	805회	60.4%	967회	58.5%
(2) 영형태	342회	25.7%	419회	25.3%
(3) ‘however’	66회	4.9%	112회	6.8%

위의 표에서 상대 빈도를 보면 55만 어절에서 70만 어절로 늘어나면서 유형 (2)는 거의 그대로인 반면, 유형 (1)과 (3)에서 약간의 증감이 일어났음을 확인할 수 있다.

3.3.3.2. ‘하지만’

다음으로 121건의 용례가 조사된 ‘하지만’의 경우를 살펴본다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 39건(9-3, 10-36), ② 접속사 ‘but’를 번역한 용례가 68건(9-19, 10-49), ③ 접속사 ‘however’를 번역한 용례가 1건(9-1, 10-0), ④ 접속사 ‘yet’를 번역한

용례가 4건(9-4, 10-0), ⑤ 접속사 ‘and’를 번역한 용례가 3건(9-0, 10-3), ⑥ 접속사 ‘though’를 번역한 용례가 2건(9-0, 10-2) ⑦ 문장부호를 번역한 용례가 0건 등으로 나타났다. 이 조사 결과를 누적 55만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

(표 25) ‘하지만’의 번역 연원 (55만 어절과 70만 어절 비교)

유형	누적 55만 어절		누적 70만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) ‘but’	594회	62.1%	662회	61.4%
(2) 영형태	259회	27.1%	298회	27.6%
(3) ‘though’	24회	2.5%	26회	2.4%
(4) ‘and’	40회	4.2%	43회	4.0%

위의 표에서 상대 빈도를 보면 55만 어절에서 70만 어절로 늘어났지만 각각의 유형에서 거의 비슷한 비율을 보이고 있음을 확인할 수 있다.

3.3.3.3. ‘그런데’

마지막으로 72건의 용례가 조사된 ‘하지만’의 경우를 살펴본다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 32건(9-11, 10-21), ② 접속사 ‘but’를 번역한 용례가 12건(9-6, 10-6), ③ 접속사 ‘however’를 번역한 용례가 1건(9-1, 10-0), ④ 접속사 ‘yet’를 번역한 용례가 1건(9-1, 10-0), ⑤ 접속사 ‘and’를 번역한 용례가 1건(9-0, 10-0), ⑥ 접속사 ‘though’를 번역한 용례가 0건 ⑦ 문장부호를 번역한 용례가 0건 등으로 나타났다. 이 조사 결과를 누적 55만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.²⁰⁾

20) 이 밖에 부사 ‘then’을 번역한 용례가 11건, 관용구 ‘by the way’를 번역한 용례가 3건, 원문과 무관하게 윤색하면서 들어간 용례가 9건 등으로 조사되었다. 이번 조사에서 3위로 나타난 ‘then’의 경우는 앞선 55만 어절까지의 조사에서 전혀 나오지 않았으므로 비율로 보면 전체적인 조사 결과를 위협할 만한 수치가 못 된다.

[표 26] '그런데'의 연원 (55만 어절과 70만 어절 비교)

유형	누적 55만 어절		누적 70만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) 영형태	187회	52.7%	219회	51.3%
(2) 'but'	84회	23.7%	96회	22.5%
(3) 'and'	60회	16.9%	63회	14.9%

위의 표에서 상대 빈도를 보면 55만 어절에서 70만 어절로 늘어나면서 각각의 유형이 모두 소폭 감소한 사실을 확인할 수 있다.

3.3.4. 줄표

새로 구축한 3차 15만 어절의 영어 원문 코퍼스에서 문장부호 줄표가 포함된 문장 용례는 모두 758건이었다. 영어 원문의 줄표는 한국어 번역문에 대략 일곱 가지 유형으로 나타났다. ① 줄표를 전후로 문장이 분리되어 새로운 문장이 생성되는 용례가 137건(9-82, 10-55), ② 줄표를 없애고 어순을 재배치한 용례가 98건(9-66, 10-32), ③ 줄표를 없애고 어순을 그대로 둔 채 풀어쓴 용례가 135건(9-103, 10-32), ④ 줄표가 있는 어구를 번역자가 누락한 용례가 24건(9-6, 10-18), ⑤ 말줄임표로 변환한 용례가 54건(9-0, 10-54), ⑥ 괄호로 처리한 용례가 18건(9-18, 10-0), ⑦ 쉼표로 대체한 용례가 0건, ⑧ 줄표로 변환한 용례가 1건(9-1, 10-0)²¹⁾ 등으로 나타났다. 이 조사 결과를 누적 55만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

21) 문장의 말미에서 줄표와 느낌표를 사용한 용례(253-9-259)인데, 여기에 포함시켰다. (ST) "Having a wonderful time—day and night included."(253-9-259) (TT) "멋진 시간을 보내며!—밤과 낮 언제나!"

(표 27) 줄표의 번역 양상 (55만 어절과 70만 어절 비교)

유형	누적 55만 어절		누적 70만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
① 새 문장	1244회	43.2%	1381회	41.2%
② 어순 도치	553회	19.2%	651회	19.4%
③ 어순 유지	586회	20.3%	721회	21.5%
④ 생략/누락	251회	8.7%	275회	8.2%
⑤ 말줄임표	75회	2.6%	129회	3.9%
⑥ 괄호	151회	5.2%	169회	5.1%
⑦ 쉼표	18회	0.6%	18회	0.5%
⑧ 줄표	5회	0.2%	6회	0.2%

위의 표를 보면, 55만 어절에서 70만 어절로 늘어나면서, 유형 ①이 소폭 감소한 반면, 유형 ③과 ⑤가 소폭 증가했음을 확인할 수 있다. (원문의 줄표를 그대로 번역문에 가져온 유형 ⑧이 이번에 1건 조사되었다.)

3.4. 누적 85만 어절

4차로 추가된 15만 어절에 대한 분석을, 기존의 70만 어절 분석에 누적시킨 결과를 항목별로 정리하면 다음과 같다.

<코퍼스 목록>

11. 『한니발 라이징』, 488쪽, 52,000어절, *Hannibal Rising*, 창해, 2007년.
12. 『성과 영혼』, 187쪽, 52,000어절, *Sex and Spirit*, 창해, 2004년.
13. 『신성한 건축』, 185쪽, 46,000어절, *The Sacred Architecture*, 창해, 2005년.

3.4.1. 의존명사 ‘때문’

새로운 4차 15만 어절 코퍼스에서 추출한 ‘때문’의 용례는 모두 258건 (11-49, 12-125, 13-84)이었다. 이들의 분포는 (1) ‘-기 때문.’ 구성이 184건 (11-21건, 12-102, 13-61), (2) ‘명사/대명사+때문’ 구성이 67건(11-27건, 12-19,

13-21), (3) ‘관형사+때문’ 구성이 4건(11-1, 12-2, 13-1), (4) 문두의 단독형 ‘때문’ 구성이 2건(11-0, 12-2, 13-0) 등으로 나타났다.²²⁾ 다시 이 수치를 앞선 70만 어절에 대한 조사 결과와 합산하면, (1) ‘-기 때문-’ 구성이 1,335건(1190+145건), (2) ‘명사/대명사+때문’ 구성이 390건(288+102), (3) ‘관형사+때문’ 구성이 256건(130+126), (4) 문두(文頭)의 단독형 ‘때문’ 구성이 25건(17+8)이 된다.

[표 28] ‘때문’의 분포 (70만 어절과 85만 어절 비교)

유형	누적 70만 어절		누적 85만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1)	1335회	66.3%	1519회	66.8%
(2)	390회	19.4%	457회	20.1%
(3)	256회	12.7%	260회	11.4%
(4)	25회	1.2%	27회	1.2%

위의 표에서 상대 빈도를 보면 70만 어절에서 85만 어절로 늘어나면서 유형 (1) ‘-기 때문-’ 구성과 (2) ‘명사/대명사+때문’ 구성은 약간 증가한 반면, 유형 (3) ‘관형사 + 때문-’ 구성은 그만큼 감소했음을 확인할 수 있다.

새로운 4차 15만 어절 병렬코퍼스에서 원문과의 대조를 통해 ‘때문’의 번역 연원을 조사한 결과는 다음과 같았다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 68건(11-14, 12-23, 13-31), ② 접속사를 번역한 용례가 96건(11-12, 12-62, 13-22), ③ 전치사를 번역한 용례가 47건(11-16, 12-14, 13-17), ④ 특정 구문을 번역한 용례가 34건(11-4, 12-20, 13-10), ⑤ 문장부호를 번역한 용례가 8건(11-1, 11-4, 12-3), ⑥ 명사를 번역한 용례가 3건(11-1, 12-0, 13-1), ⑦ 부사를 번역한 용례가 2건(11-1, 12-0, 13-1) 등으로 나타났다. 이 조사 결과를 누적 70만 어절에 대한 조

22) 각주 17)에서와 같이 ‘용언의 관형형+때문’ 유형이 이번에도 1건 조사되었다. (예) 중앙아메리카의 신전은 대부분 전면을 강조하고 있는데, 이는 외부로부터의 시선을 중시한 때문이었다. (60-13-113) 지금까지 누적 85만 어절을 기준으로 0.37(반올림 0.4%)에 해당하는 분포이다.

사 결과와 합산해서 표로 만들면 아래와 같다.

[표 29] '때문'의 번역 연원 (70만 어절과 85만 어절 비교)

유형	누적 70만 어절		누적 85만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
① 영형태	620회	30.8%	688회	30.3%
② 접속사	889회	44.1%	985회	43.3%
③ 전치사	193회	9.6%	240회	10.6%
④ 특정 구문	239회	11.9%	273회	12.1%
⑤ 문장부호	41회	2.0%	49회	2.1%
⑥ 명사/부사	32회	1.6%	37회	1.6%

위의 표에서 상대 빈도를 보면 70만 어절에서 85만 어절로 늘어나면서 유형 ②와 ③의 비율에서 다소간의 증감을 보인 사실을 확인할 수 있다.

3.4.2. 장형사동 구문

새로운 4차 15만 어절 코퍼스에서 추출한 장형사동 구문의 용례는 모두 204건이었다. 이들 용례 중에서 장형사동을 단형사동으로 교체할 수 있는 경우는 17건(8.3%)으로 나타났는데, 이 수치를 앞의 누적 70만 어절에 대한 조사 결과와 합산하면 모두 92건(6.6%, 92/1404)이 된다. 이는 누적 70만 어절을 대상으로 한 조사에서 얻었던 결과와 거의 비슷한 수치이다

[표 30] 단형사동 교체 가능 용례 (70만 어절과 85만 어절 비교)

유형	누적 70만 어절		누적 85만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
장형→ 단형	75회	6.3%	92건	6.6%

새로운 4차 15만 어절 코퍼스에서 원문과의 대조를 통해 장형사동 구문의 번역 연원을 조사한 결과는 다음과 같았다. ① 사역동사 'have'를 번역한 용례

가 0건, ② 사역동사 ‘make’를 번역한 용례가 27건(11-2, 12-19, 13-6), ③ 사역동사 ‘let’을 번역한 용례가 5건(11-3, 12-2, 13-0), ④ 사역동사 ‘get’을 번역한 용례가 4건(11-0, 12-0, 13-4), ⑤ 사역의 의미자질을 가진 일반동사를 번역한 용례가 10건(11-1, 12-9, 13-0), ⑥ 일반동사를 번역한 용례가 106건(11-17, 12-69, 13-20), ⑦ 의역에서 비롯된 용례가 52건(11-15, 12-28, 13-9) 등으로 나타났다. 이 조사 결과를 누적 70만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

(표 31) 장형사동 구문의 번역 연원 (70만 어절과 85만 어절 비교)

유형	누적 70만 어절		누적 85만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) 사역동사	205회	17.1%	244회	17.3%
(2) 일반동사	595회	49.6%	711회	50.6%
(3) 의역	400회	33.3%	452회	32.1%

위의 표에서 상대 빈도를 보면 70만 어절에서 85만 어절로 늘어나면서 유형의 순서는 그대로이지만 유형 (2)와 (3)에서 소폭의 증감이 있음을 확인할 수 있다.

3.4.3. 접속부사

새로운 4차 15만 어절 코퍼스에서 추출한 접속부사 ‘그러나’, ‘하지만’, ‘그런데’ 등의 용례는 모두 344건이었다. 이들 용례를 직접 대조(‘그러나’와 ‘하지만’)와 간접 대조(‘그런데’)로 나누어 그 출현 횟수를 비교해보면 329건(208+121)과 15건으로 나타났다. 이 수치를 앞의 누적 70만 어절에 대한 조사 결과와 합산하면, 각각 3,061건(2732+329)과 442건(427+15)으로 나타났다.

(표 32) 직접 대조와 간접 대조의 빈도 (70만 어절과 85만 어절 비교)

유형	누적 70만 어절		누적 85만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
간접/직접	427/2732	15.6%	442/3061	14.4%

위의 도표를 보면 70만 어절에서 85만 어절로 늘어나면서 직접 대조와 간접 대조 용법의 비율이 소폭 감소했음을 알 수 있다.

새로운 4차 15만 어절 병렬코퍼스에서 원문과의 대조를 통해 ‘그러나’와 ‘하지만’ 및 ‘그런데’의 번역 연원을 조사한 결과는 다음과 같았다.

3.4.3.1. ‘그러나’

먼저 모두 208건의 용례가 조사된 ‘그러나’의 경우를 살펴본다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 43건(11-13, 12-9, 13-21), ② 접속사 ‘but’를 번역한 용례가 83건(11-12, 12-37, 13-34), ③ 접속사 ‘however’를 번역한 용례가 54건(11-0, 12-47, 13-7), ④ 접속사 ‘yet’를 번역한 용례가 14건(11-0, 12-11, 12-3), ⑤ 접속사 ‘and’를 번역한 용례가 7건(11-3, 12-2, 1-2), ⑥ 접속사 ‘though’를 번역한 용례가 6건(11-0, 12-5, 13-1), ⑦ 문장부호를 번역한 용례가 1건(11-0, 12-1, 13-0) 등으로 나타났다. 이 조사 결과를 70만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

(표 33) ‘그러나’의 번역 연원 (70만 어절과 85만 어절 비교)

유형	누적 70만 어절		누적 85만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) ‘but’	967회	58.5%	1050회	56.4%
(2) 영형태	419회	25.3%	462회	24.8%
(3) ‘however’	112회	6.8%	166회	8.9%

위의 표에서 상대 빈도를 보면 70만 어절에서 85만 어절로 늘어나면서 유형 (2)는 거의 그대로인 반면, 유형 (1)과 (3)에서 2% 내외의 증감이 일어난 사실을 확인할 수 있다.

3.4.3.2. ‘하지만’

다음으로 121건의 용례가 조사된 ‘하지만’의 경우를 살펴본다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용

례가 64건(11-59, 12-4, 13-1), ② 접속사 ‘but’를 번역한 용례가 50건(11-45, 12-5, 13-0), ③ 접속사 ‘however’를 번역한 용례가 1건(11-0, 12-1, 13-0), ④ 접속사 ‘yet’를 번역한 용례가 0건, ⑤ 접속사 ‘and’를 번역한 용례가 3건(11-2, 12-0, 13-1), ⑥ 접속사 ‘though’를 번역한 용례가 3건(11-0, 12-3, 13-0) ⑦ 문장부호를 번역한 용례가 0건 등으로 나타났다. 이 조사 결과를 누적 70만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

(표 34) ‘하지만’의 번역 연원 (70만 어절과 85만 어절 비교)

유형	누적 70만 어절		누적 85만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) ‘but’	662회	61.4%	712회	59.4%
(2) 영형태	298회	27.6%	362회	30.2%
(3) ‘though’	26회	2.4%	29회	2.4%
(4) ‘and’	43회	4.0%	46회	4.2%

위의 표에서 상대 빈도를 보면 70만 어절에서 85만 어절로 늘어나면서 유형 (1)과 (2)에서 2% 내외의 증감이 있음을 확인할 수 있다.

3.4.3.3. ‘그런데’

마지막으로 15건의 용례가 조사된 ‘하지만’의 경우를 살펴본다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 8건(11-4, 12-4, 13-0), ② 접속사 ‘but’를 번역한 용례가 0건, ③ 접속사 ‘however’를 번역한 용례가 0건, ④ 접속사 ‘yet’를 번역한 용례가 2건(11-0, 12-0, 13-2), ⑤ 접속사 ‘and’를 번역한 용례가 5건(11-4, 12-1, 13-0), ⑥ 접속사 ‘though’를 번역한 용례가 0건 ⑦ 문장부호를 번역한 용례가 0건 등으로 나타났다. 이 조사 결과를 누적 70만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

(표 35) '그런데'의 번역 연원 (70만 어절과 85만 어절 비교)

유형	누적 70만 어절		누적 85만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) 영형태	219회	51.3%	227회	51.4%
(2) 'but'	96회	22.5%	96회	21.7%
(3) 'and'	63회	14.9%	68회	14.3%

위의 표에서 상대 빈도를 보면 70만 어절에서 85만 어절로 늘어나면서 유형 (2)와 (3)이 소폭 감소했음을 확인할 수 있다.

3.4.4. 줄표

새로 구축한 4차 15만 어절의 영어 원문 코퍼스에서 문장부호 줄표가 포함된 문장 용례는 모두 351건이었다. 영어 원문의 줄표는 한국어 번역문에 대략 일곱 가지 유형으로 나타났다. ① 줄표를 전후로 문장이 분리되어 새로운 문장이 생성되는 용례가 103건(11-50, 12-22, 13-31), ② 줄표를 없애고 어순을 재배치한 용례가 89건(11-3, 12-53, 13-33), ③ 줄표를 없애고 어순을 그대로 둔 채 풀어쓴 용례가 117건(11-9, 12-61, 13-47), ④ 줄표가 있는 어구를 번역자가 누락한 용례가 13건(11-4, 12-6, 13-3), ⑤ 말줄임표로 변환한 용례가 22건(11-22, 12-0, 13-0), ⑥ 괄호로 처리한 용례가 5건(11-0, 12-3, 13-2), ⑦ 쉼표로 대체한 용례가 0건, ⑧ 줄표로 변환한 용례가 2건(11-22³⁾, 12-0, 13-0) 등으로 나타났다. 이 조사 결과를 누적 70만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

(표 36) 줄표의 번역 양상 (70만 어절과 85만 어절 비교)

유형	누적 70만 어절		누적 85만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
① 새 문장	1381회	41.2%	1484회	40.1%

23) 문장의 말미에 말줄임표 대신 사용한 것과 문장 중간에 연관 부호로 사용한 것이 조사되었다. (예) “미사의 목소리-”(56-11-198), ‘포도주 저장용-서늘한 장소에 보관하시오’(80-11-268).

② 어순 도치	651회	19.4%	740회	20.0%
③ 어순 유지	721회	21.5%	838회	22.6%
④ 생략/누락	275회	8.2%	288회	7.8%
⑤ 말줄임표	129회	3.9%	151회	4.1%
⑥ 괄호	169회	5.1%	174회	4.7%
⑦ 쉼표	18회	0.5%	18회	0.5%
⑧ 줄표	6회	0.2%	8회	0.2%

위의 표를 보면, 70만 어절에서 85만 어절로 늘어나면서, 유형 ①의 비율이 1% 내외로 감소하고 다른 유형은 거의 비슷한 비율을 유지하고 있음을 확인할 수 있다.

3.5. 누적 100만 어절

5차로 추가된 15만 어절에 대한 분석을, 기존의 85만 어절 분석에 누적시킨 결과는 다음과 같다.

<코퍼스 목록>²⁴⁾

14. 『론리플래닛 스토리』, 476쪽, 83,000어절, *The Lonely Planet Story*, 안 그래픽스, 2008년.
15. 『신성한 지구』, 169쪽, 67,000어절, *The Sacred Earth*, 창해, 2002년.

3.5.1. 의존명사 ‘때문’

새로운 5차 15만 어절 코퍼스에서 추출한 ‘때문’의 용례는 모두 296건(14-217, 15-79)이었다. 이들의 분포는 (1) ‘-기 때문-’ 구성이 192건(14-146건, 15-46), (2) ‘명사/대명사+때문’ 구성이 91건(14-61건, 15-30), (3) ‘관형사+때문’ 구성이 7건(14-5, 15-2), (4) 문두의 단독형 ‘때문’ 구성이 6건(14-5, 15-1) 등으로 나타났다. 다시 이 수치를 앞선 85만 어절에 대한 조사 결과와 합산하면, (1) ‘-기 때문-’ 구성이 1,711건(1519+192건), (2) ‘명사/대명사+때문’ 구성이 548건

24) 원래 자료 14의 코퍼스 크기는 97,000어절이었는데, 양적 변수의 크기를 일정하게 맞추기 위해 15% 정도를 감축해서 83,000어절까지만 조사했다.

(457+91), (3) ‘관형사+때문’ 구성이 267건(260+7), (4) 문두(文頭)의 단독형 ‘때문’ 구성이 33건(27+6)이 된다.

[표 37] ‘때문’의 분포 (85만 어절과 100만 어절 비교)

유형	누적 85만 어절		누적 100만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1)	1519회	66.8%	1711회	66.6%
(2)	457회	20.1%	548회	21.3%
(3)	260회	11.4%	267회	10.4%
(4)	27회	1.2%	33회	1.3%

위의 표에서 상대 빈도를 보면 85만 어절에서 100만 어절로 늘어나면서 유형 (2) ‘명사/대명사+때문-’ 구성과 유형 (3) ‘관형사 + 때문-’ 구성에서 1% 내외의 증감이 일어났음을 확인할 수 있다.

새로운 5차 15만 어절 병렬코퍼스에서 원문과의 대조를 통해 ‘때문’의 번역 연원을 조사한 결과는 다음과 같았다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 94건(14-78, 15-16), ② 접속사를 번역한 용례가 110건(14-81, 15-29), ③ 전치사를 번역한 용례가 40건(14-23, 15-17), ④ 특정 구문을 번역한 용례가 36건(11-23, 15-13), ⑤ 문장부호를 번역한 용례가 13건(14-9, 15-4), ⑥ 명사를 번역한 용례가 3건(14-3, 15-0), ⑦ 부사를 번역한 용례가 0건 등으로 나타났다. 이 조사 결과를 누적 85만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

[표 38] ‘때문’의 번역 연원 (85만 어절과 100만 어절 비교)

유형	누적 85만 어절		누적 100만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
① 영형태	688회	30.3%	782회	30.4%
② 접속사	985회	43.3%	1095회	42.6%
③ 전치사	240회	10.6%	280회	11.0%

④ 특정 구문	273회	12.1%	309회	12.0%
⑤ 문장부호	49회	2.1%	62회	2.4%
⑥ 명사/부사	37회	1.6%	40회	1.6%

위의 표에서 상대 빈도를 보면 85만 어절에서 100만 어절로 늘어나면서 유형 ②와 ③의 비율에서 다소간의 증감을 보인 사실을 확인할 수 있다.

3.5.2. 장형사동 구문

새로운 5차 15만 어절 코퍼스에서 추출한 장형사동 구문의 용례는 모두 71건이었다. 이들 용례 중에서 장형사동을 단형사동으로 교체할 수 있는 경우는 6건(8.5%)으로 나타났는데, 이 수치를 앞의 누적 85만 어절에 대한 조사 결과와 합산하면 모두 98건(6.6%, 98/1475)이 된다. 이는 누적 85만 어절을 대상으로 한 조사에서 얻었던 결과와 동일한 수치이다

(표 39) 단형사동 교체 가능 용례 (85만 어절과 100만 어절 비교)

유형	누적 85만 어절		누적 100만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
장형→ 단형	92회	6.6%	98건	6.6%

새로운 5차 15만 어절 코퍼스에서 원문과의 대조를 통해 장형사동 구문의 연원을 조사한 결과는 다음과 같았다. ① 사역동사 ‘have’를 번역한 용례가 0건, ② 사역동사 ‘make’를 번역한 용례가 6건(14-3, 15-3), ③ 사역동사 ‘let’을 번역한 용례가 2건(14-2, 15-0), ④ 사역동사 ‘get’을 번역한 용례가 2건(14-2, 15-0), ⑤ 사역의 의미자질을 가진 일반동사를 번역한 용례가 4건(14-1, 15-3), ⑥ 일반동사를 번역한 용례가 28건(14-18, 15-10), ⑦ 의역에서 비롯된 용례가 29건(14-14, 15-15) 등으로 나타났다. 이 조사 결과를 누적 85만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

(표 40) 장형사동 구문의 번역 연원 (85만 어절과 100만 어절 비교)

유형	누적 70만 어절		누적 85만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) 사역동사	244회	17.4%	254회	17.1%
(2) 일반동사	711회	50.6%	743회	50.3%
(3) 의역	452회	32.1%	481회	32.6%

위의 표에서 상대 빈도를 보면 85만 어절에서 100만 어절로 늘어났지만 유형의 순서와 상대 빈도는 엇비슷하게 나타났음을 확인할 수 있다.

3.5.3. 접속부사

새로운 5차 15만 어절 코퍼스에서 추출한 접속부사 ‘그러나’, ‘하지만’, ‘그런데’ 등의 용례는 모두 408건이었다. 이들 용례를 직접 대조(‘그러나’와 ‘하지만’)와 간접 대조(‘그런데’)로 나누어 그 출현 횟수를 비교해보면 318건(157+161)과 90건으로 나타났다. 이 수치를 앞의 누적 85만 어절에 대한 조사 결과와 합산하면, 각각 3,540건(3050+490)과 532건(442+90)으로 나타났다.

(표 41) 직접 대조와 간접 대조의 빈도 (85만 어절과 100만 어절 비교)

유형	누적 85만 어절		누적 100만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
간접/직접	442/2732	14.4%	532/3540	15.2%

위의 도표를 보면 85만 어절에서 100만 어절로 늘어나면서 직접 대조와 간접 대조 용법의 비율이 소폭 증가했음을 알 수 있다.

새로운 5차 15만 어절 병렬코퍼스에서 원문과의 대조를 통해 ‘그러나’와 ‘하지만’ 및 ‘그런데’의 번역 연원을 조사한 결과는 다음과 같았다.

3.5.3.1. ‘그러나’

먼저 모두 157건의 용례가 조사된 ‘그러나’의 경우를 살펴본다. ① (원문에

대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 57건(14-55, 15-2), ② 접속사 ‘but’를 번역한 용례가 67건(14-61, 15-6), ③ 접속사 ‘however’를 번역한 용례가 26건(14-8, 15-18), ④ 접속사 ‘yet’를 번역한 용례가 0건, ⑤ 접속사 ‘and’를 번역한 용례가 2건(14-2, 15-0), ⑥ 접속사 ‘though’를 번역한 용례가 5건(14-5, 15-0), ⑦ 문장부호를 번역한 용례가 0건 등으로 나타났다. 이 조사 결과를 85만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

[표 42] ‘그러나’의 번역 연원 (85만 어절과 100만 어절 비교)

유형	누적 85만 어절		누적 100만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) ‘but’	1050회	56.4%	1117회	55.3%
(2) 영형태	462회	24.8%	519회	25.7%
(3) ‘however’	166회	8.9%	192회	8.2%

위의 표에서 상대 빈도를 보면 85만 어절에서 100만 어절로 늘어나면서 유형 (1)과 (2), (3) 모두의 비율에서 약간의 증감이 있음을 확인할 수 있다.

3.5.3.2. ‘하지만’

다음으로 161건의 용례가 조사된 ‘하지만’의 경우를 살펴본다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 58건(14-56, 15-2), ② 접속사 ‘but’를 번역한 용례가 78건(14-74, 15-4), ③ 접속사 ‘however’를 번역한 용례가 18건(14-6, 15-13), ④ 접속사 ‘yet’를 번역한 용례가 0건, ⑤ 접속사 ‘and’를 번역한 용례가 1건(14-1, 15-0), ⑥ 접속사 ‘though’를 번역한 용례가 2건(14-2, 15-0) ⑦ 문장부호를 번역한 용례가 3건(14-2, 15-1) 등으로 나타났다. 이 조사 결과를 누적 85만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

[표 43] '하지만'의 번역 연원 (85만 어절과 100만 어절 비교)

유형	누적 85만 어절		누적 100만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) 'but'	712회	59.4%	790회	58.1%
(2) 영형태	362회	30.2%	420회	30.1%
(3) 'though'	29회	2.4%	31회	2.3%
(4) 'and'	46회	4.2%	47회	3.5%

위의 표에서 상대 빈도를 보면 85만 어절에서 100만 어절로 늘어나면서 유형 (1)에서 1% 남짓한 감소가 일어났고 (4)에서 1% 미만의 감소가 일어났음을 확인할 수 있다.

3.5.3.3. '그런데'

마지막으로 90건의 용례가 조사된 '하지만'의 경우를 살펴본다. ① (원문에 대응하는 어구가 없는 상태에서 문맥에서 유추해 도입한) 영형태를 번역한 용례가 70건(14-69, 15-1), ② 접속사 'but'를 번역한 용례가 14건(14-14, 15-0), ③ 접속사 'however'를 번역한 용례가 0건, ④ 접속사 'yet'를 번역한 용례가 0건, ⑤ 접속사 'and'를 번역한 용례가 4건(14-4, 15-0), ⑥ 접속사 'though'를 번역한 용례가 1건(14-1, 15-0) ⑦ 문장부호를 번역한 용례가 1건(14-1, 15-0) 등으로 나타났다. 이 조사 결과를 누적 85만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

[표 44] '그런데'의 번역 연원 (85만 어절과 100만 어절 비교)

유형	누적 85만 어절		누적 100만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
(1) 영형태	227회	51.4%	297회	55.8%
(2) 'but'	96회	21.7%	110회	20.7%
(3) 'and'	68회	14.3%	72회	13.5%

위의 표에서 상대 빈도를 보면 85만 어절에서 100만 어절로 늘어나면서 유

형 (1)이 4% 남짓 증가하고 유형 (2)와 (3)이 1% 내외로 감소했음을 확인할 수 있다.

3.5.4. 줄표

새로 구축한 5차 15만 어절의 영어 원문 코퍼스에서 문장부호 줄표가 포함된 문장 용례는 모두 239건이었다. 영어 원문의 줄표는 한국어 번역문에 대략 일곱 가지 유형으로 나타났다. ① 줄표를 전후로 문장이 분리되어 새로운 문장이 생성되는 용례가 95건(14-86, 15-9), ② 줄표를 없애고 어순을 재배치한 용례가 49건(14-29, 15-20), ③ 줄표를 없애고 어순을 그대로 둔 채 풀어쓴 용례가 56건(14-42, 15-14), ④ 줄표가 있는 어구를 번역자가 누락한 용례가 21건(14-21, 15-0), ⑤ 말줄임표로 변환한 용례가 0건, ⑥ 괄호로 처리한 용례가 18건(14-15, 15-0), ⑦ 쉼표로 대체한 용례가 0건, ⑧ 줄표로 변환한 용례가 0건 등으로 나타났다. 이 조사 결과를 누적 85만 어절에 대한 조사 결과와 합산해서 표로 만들면 아래와 같다.

[표 45] 줄표의 번역 양상 (85만 어절과 100만 어절 비교)

유형	누적 85만 어절		누적 100만 어절	
	출현 횟수	상대 빈도	출현 횟수	상대 빈도
① 새 문장	1484회	40.1%	1579회	40.1%
② 어순 도치	740회	20.0%	789회	20.0%
③ 어순 유지	838회	22.6%	894회	22.7%
④ 생략/누락	288회	7.8%	309회	7.8%
⑤ 말줄임표	151회	4.1%	151회	3.8%
⑥ 괄호	174회	4.7%	192회	4.9%
⑦ 쉼표	18회	0.5%	18회	0.5%
⑧ 줄표	8회	0.2%	8회	0.2%

위의 표를 보면, 85만 어절에서 100만 어절로 늘어나면서, 유형 ⑤와 ⑥에서 약간의 증감이 있음을 확인할 수 있다.

4. 양적 인자의 증가와 분석 결과의 변화 추이

우리는 앞선 제3장에서 코퍼스의 크기를 15만 어절씩 누적시켜 가면서 앞서 25만 어절의 조사 결과가 어떻게 변화하는가를 5회에 걸쳐 관찰했다. 여기서는 양적 인자의 변화에 따른 조사 결과의 변화 추이를 항목별로 정리해 보기로 한다.

4.1. 의존명사 '때문'

앞서도 언급했듯이 우리의 관심사는 코퍼스의 규모에 따른 분석 결과의 변화 양상이다. 즉, 코퍼스의 크기가 어느 규모 이상이면 그로부터 도출되는 분석 결과의 일반성을 담보할 수 있는냐가 문제의 핵심이다. 이러한 맥락에서 우리는 어절 수 표시 구간의 백분율 증감에 주목했다.

[표 46] '때문' 분포의 상대 빈도 변화 추이

유형/어절	누적 어절 수					
	25만	40만	55만	70만	85만	100만
(1)	74.6%	73.5%	72.9%	66.3%	66.8%	66.6%
(2)	23.4%	20.7%	17.6%	19.4%	20.1%	21.3%
(3)	1.8%	5.6%	8.0%	12.7%	11.4%	10.4%
(4)	0.2%	0.2%	1.0%	1.2%	1.2%	1.3%

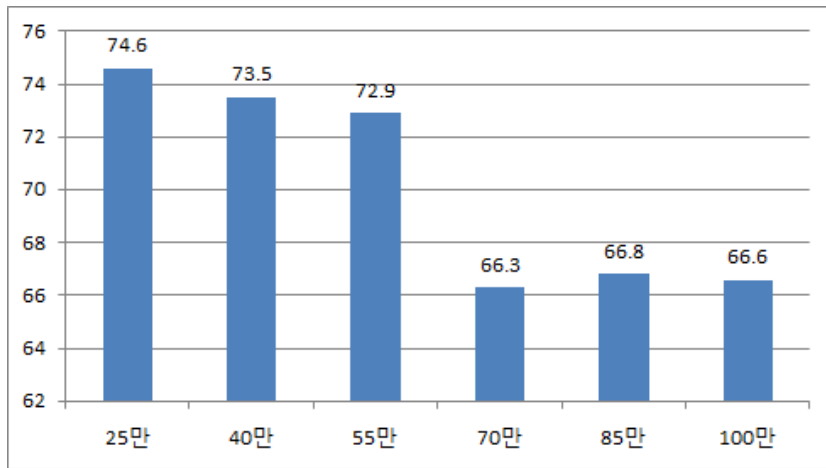
위의 [표 46]에서 유형 (1)을 보면 55만 어절에서 70만 어절로 늘어나는 구간에서 급격한 변화를 보이고, 그 이전과 이후로는 변화의 정도가 미미함을 알 수 있다. 이것을 알아보기 쉽게 그래프로 그리면 다음의 [표 47]과 같다.

이 그래프를 보면 급격한 변화 구간이 어디인지 쉽게 알아낼 수 있지만, 우리는 좀더 객관성을 확보하기 위해 증감 변화량의 표준편차(SD)를 구해 $2*SD$ 값을 넘거나 혹은 그 값에 근접하는 값을 유의미한 변화로 간주하기로 했다.²⁵⁾

25) 통계학에서는 이러한 경우에 표준편차의 두 배가 넘는 변량은 우연으로 나타날 수 없는, 유의미한 것으로 본다고 한다.

예컨대 ‘때문’의 분포 4개 유형을 구간별 변화량에 초점을 맞추어 다시 표로 만들고 이를 근거로 얻은 기술통계 결과는 다음의 [표 48] 및 [표 49]와 같다.

[표 47] ‘때문’ 분포 유형 (1) (‘-기 때문-’ 구성)의 막대그래프



[표 48] ‘때문’ 분포 유형의 구간별 변화표

유형 / 증감	25만-40만	40만-55만	55만-70만	70만-85만	85만-100만
(1)	-1.1	-0.6	-6.6	+0.5	-0.2
(2)	-2.7	-3.1	+1.8	+0.7	+1.2
(3)	+3.8	+2.4	+4.7	-1.3	-1.0
(4)	0.0	+0.8	+0.2	0.0	+0.1

[표 49] ‘때문’ 분포 유형의 변화량에 따른 기술통계량

	표본 수효	최솟값	최댓값	평균	표준편차
유형 (1)	5	-6.60	+0.50	-1.60	2.8557
유형 (2)	5	-3.10	+2.70	+0.46	2.2142
유형 (3)	5	-1.30	+4.70	+1.72	2.7471
유형 (4)	5	0	+0.80	+0.22	0.3346

위의 [표 49]에서 유형 (1)의 표준편차를 구하면 2.8557이 되어, 2*SD 값은 ± 5.7114 이다. 우리는 [표 48]의 55만-70만 어절 구간에서 2*SD 값을 넘는 변화량 -6.6%를 확인할 수 있다. 이러한 변화는 앞서 언급했듯이 유의미한 변화이며, 따라서 이 구간을 유의미한 코퍼스 크기를 결정하는 기준으로 볼 수 있다. 즉, ‘때문’의 분포 유형 (1)은 코퍼스의 크기가 70만 어절이 되면 그로부터 도출된 분석 결과를 신뢰할 수 있다는 것이다. 나머지 분포 유형에 대해서도 동일한 방식으로 변화량과 표준편차를 구하여, 이를 근거로 각각의 유형에서 유의미한 변화를 보이는 구간을 추출했다. 그 결과, 유형 (2)에서 40만-55만 구간 (-3.10), 유형 (3)에서 55만-70만 구간(+4.70), 유형 (4)에서 55만-70만 구간 (-0.80)을 각각 유의미한 코퍼스 규모를 결정하는 기준으로 확인할 수 있었다.²⁶⁾ 위의 4개 유형에서 나온 결과를 종합하면, ‘때문’의 분포 유형은 70만 어절이 확보되어야 그 상대 빈도를 성공적으로 예측할 수 있다는 결론을 얻을 수 있다.

다음으로 우리는 아래 [표 50]에 근거하여 ‘때문’의 번역 연원에 대해서도 동일한 방식으로 어절수의 증가에 따른 6개 유형의 변화량과 그 표준편차를 계산해 보았다.²⁷⁾

[표 50] ‘때문’ 번역 연원의 상대 빈도 변화 추이

유형	누적 어절 수					
	25만	40만	55만	70만	85만	100만
① 영형태	44.5%	34.9%	32.8%	30.8%	30.3%	30.4%
② 접속사	39.9%	42.6%	43.7%	44.1%	43.3%	42.6%
③ 전치사	7.2%	8.4%	7.4%	9.6%	10.6%	11.0%
④ 구문	3.4%	10.1%	12.3%	11.9%	12.1%	12.0%
⑤ 부호	3.0%	2.2%	2.1%	2.0%	2.1%	2.4%
⑥ N/Ad	2.0%	1.8%	1.7%	1.6%	1.6%	1.6%

26) 유형 (2)는 $2.2142 < (-)3.1 < 4.4284$ 이고, (3)은 $2.7471 < 4.7 < 5.4942$ 가 된다. 즉, 이들 값은 2*SD를 넘지 못하지만, 1*SD를 넘는 2*SD의 근사값이다.

27) 분석과 논의의 검증을 위해 변화량의 증감 일람표와 이들에 대한 표준편차 일람표를 논문 말미에 부록으로 제시했다.

그 결과, 유형 ③을 제외한 모든 유형에서 25만-40만 어절 구간이 유의미한 코퍼스의 규모를 결정하는 기준으로 나타났고, 유형 ③에서만 55만-70만 어절 구간(+2.2)이 유의미한 기준으로 나타났다. 그러므로 우리는 ‘때문’의 번역 연원에 대해 최소 40만 어절에서 최대 70만 어절이 확보되면 그 상대 빈도를 예측할 수 있다는 결론을 얻을 수 있다.

4.2. 장형사동

먼저 장형사동을 단형사동으로 교체할 수 있는 형태의 빈도 변화를 보이는 [표 51]을 기본 자료로 삼아 앞서와 동일한 방식으로 변화량의 증감표를 만들고 다시 이를 근거로 변화량의 표준편차를 구해 보았다.

[표 51] 단형사동 교체 가능 형태의 빈도 변화 추이

유형	누적 어절 수					
	25만	40만	55만	70만	85만	100만
단형 → 장형	6.0%	5.4%	5.8%	6.3%	6.6%	6.6%

그 결과, 우리는 1*SD와 2*SD의 중간값을 엇비슷하게 갖는 25만-40만 어절 구간과 55만-70만 구간을 유의미한 기준으로 추출했다.²⁸⁾ 그러므로 우리는 70만 어절 크기의 코퍼스를 확보하면 장형에서 단형으로 교체가 가능한 형태의 발생 빈도를 예측할 수 있다는 결론을 얻을 수 있다.

다음으로 우리는 아래 [표 52]에 근거하여 장형사동 구문의 번역 연원에 대해서도 동일한 방식으로 어절수의 증가에 따른 3개 유형의 변화량과 그 표준편차를 계산해 보았다.

28) 이들의 변화량은 각각 -0.6과 +0.5인데 모두 1*SD(0.44)와 2*SD(0.88)의 중간에 위치한다.

[표 52] 장형사동 번역 연원의 상대 빈도 변화 추이

유형	누적 어절 수					
	25만	40만	55만	70만	85만	100만
(1) 사역동사	14.5%	17.5%	17.5%	17.1%	17.3%	17.1%
(2) 일반동사	49.3%	49.3%	47.8%	49.6%	50.6%	50.3%
(3) 의역	36.2%	33.2%	34.7%	33.3%	32.1%	32.6%

그 결과, 우리는 사역동사 구문을 번역한 유형 (1)과 의역에 의한 유형 (3)에서 25만-40만 어절 구간이 유의미한 코퍼스의 규모를 결정하는 기준으로 나타났고, 일반동사 구문을 번역한 유형 (2)에서 55만-70만 어절 구간이 유의미한 기준으로 나타났다. 그러므로 우리는 장형사동 구문의 번역 연원에 대해 70만 어절이 확보되면 그 상대 빈도를 예측할 수 있다는 결론을 얻을 수 있다.

4.3. 문장 접속부사

먼저 직접대조(‘그러나’, ‘하지만’)와 간접대조(‘그런데’)의 비율 변화 추이를 나타내는 [표 53]을 기본 자료로 삼아 앞서와 동일한 방식으로 변화량의 증감표를 만들고 다시 이를 근거로 변화량의 표준편차를 구해 보았다.

[표 53] 직접대조와 간접대조의 빈도 변화 추이

유형	누적 어절 수					
	25만	40만	55만	70만	85만	100만
간접/직접	17.2%	12.2%	15.5%	15.6%	14.4%	15.2%

그 결과, 우리는 25만-40만 어절 구간을 유의미한 코퍼스의 규모를 결정하는 기준으로 확인했다. 그러므로 우리는 40만 어절 크기의 코퍼스를 확보하면 직접대조와 간접대조의 상대 빈도를 예측할 수 있다는 결론을 얻을 수 있다.

다음으로 우리는 아래 [표 54], [표 55], [표 56]에 근거하여 각각 ‘그러나’와 ‘하지만’ 및 ‘그런데’의 번역 연원에 대해서도 동일한 방식으로 어절수의 증가에 따른 3-4개 유형의 변화량과 그 표준편차를 계산해 보았다.

4.3.1. ‘그러나’

(표 54) ‘그러나’ 번역 연원의 상대 빈도 변화 추이

유형	누적 어절 수					
	25만	40만	55만	70만	85만	100만
(1) but	41.0%	61.9%	60.4%	58.5%	56.4%	55.3%
(2) 영형태	33.3%	21.5%	25.7%	25.3%	24.8%	25.7%
(3) however	17.1%	7.1%	4.9%	6.8%	8.9%	8.2%

조사 결과, 우리는 접속사 ‘but’을 번역한 유형 (1)과 문맥에서 유추해서 영 형태를 번역한 유형 (2) 및 접속사 ‘however’를 번역한 유형 (3)에서 모두 25만-40만 어절 구간을 유의미한 코퍼스 크기의 결정 기준으로 확인할 수 있었다. 그러므로 우리는 40만 어절이 확보되면 문장접속부사 ‘그러나’가 영어 원문의 무엇을 번역한 것인가에 대한 답변을 예측할 수 있다는 결론을 얻게 된다.

4.3.2. ‘하지만’

(표 55) ‘하지만’ 번역 연원의 상대 빈도 변화 추이

유형	누적 어절 수					
	25만	40만	55만	70만	85만	100만
(1) but	64.6%	65.2%	62.1%	61.4%	59.4%	58.1%
(2) 영형태	20.8%	22.1%	27.1%	27.6%	30.2%	30.1%
(3) though	3.5%	3.4%	2.5%	2.4%	2.4%	2.3%
(4) and	3.0%	3.2%	4.2%	4.0%	4.2%	3.5%

조사 결과, 우리는 접속사 ‘but’을 번역한 유형 (1)과 문맥에서 유추해서 영 형태를 번역한 유형 (2), 접속사 ‘though’를 번역한 유형 (3), 접속사 ‘and’를 번역한 유형 (4)에서 모두 40만-55만 어절 구간을 유의미한 코퍼스 크기의 결정 기준으로 확인할 수 있었다. 그러므로 우리는 55만 어절이 확보되면 문장접속 부사 ‘하지만’이 영어 원문의 무엇을 번역한 것인가에 대한 답변을 예측할 수 있다는 결론을 얻게 된다.

4.3.3. ‘그런데’

[표 56] ‘그런데’ 번역 연원의 상대 빈도 변화 추이

유형	누적 어절 수					
	25만	40만	55만	70만	85만	100만
(1) 영형태	45.8%	42.3%	52.7%	51.3%	51.4%	55.8%
(2) but	28.0%	30.2%	23.7%	22.5%	21.7%	20.7%
(3) and	15.2%	18.1%	16.9%	14.9%	14.3%	13.5%

조사 결과, 우리는 문맥에서 유추해서 영형태를 번역한 유형 (1)과 접속사 ‘but’을 번역한 유형 (2)에서는 40만-55만 어절 구간을, 접속사 ‘and’를 번역한 유형 (3)에서는 25만-40만 어절 구간을 각각 유의미한 코퍼스 크기의 결정 기준으로 확인할 수 있었다. 그러므로 우리는 55만 어절이 확보되면 문장접속부사 ‘그런데’가 영어 원문의 무엇을 번역한 것인가에 대한 답변을 예측할 수 있다는 결론을 얻게 된다.

4.4. 줄표

영어 원문의 줄표가 한국어 번역문에서 어떻게 번역되었는가를 조사한 결과가 코퍼스의 크기 확대에 따라 변화한 양상을 기록한 아래 [표 57]을 바탕으로 앞서와 동일한 방식으로 코퍼스의 크기에 따른 변화량 일람표를 작성한 다음에 이들 변화량의 표준편차를 구해 보았다.

[표 57] 줄표 번역 연원의 상대 빈도 변화 추이

유형	누적 어절 수					
	25만	40만	55만	70만	85만	100만
①새 문장	40.5%	39.1%	43.2%	41.2%	40.1%	40.1%
②어순도치	19.5%	20.2%	19.2%	20.0%	20.0%	20.0%
③어순유지	18.1%	21.1%	20.3%	22.6%	22.6%	22.7%
④생략/누락	11.2%	8.7%	8.7%	7.8%	7.8%	7.8%
⑤말줄임표	7.1%	3.5%	2.6%	4.1%	4.1%	3.8%

⑥ 괄호	2.3%	6.8%	5.2%	4.7%	4.7%	4.9%
⑦ 쉽표	2.3%	0.6%	0.6%	0.5%	0.5%	0.5%
⑧ 줄표	0%	0%	0.2%	0.2%	0.2%	0.2%

조사 결과, 우리는 줄표를 전후로 문장이 분리되어 새로운 문장이 생성되는 유형 ①과 줄표를 없애고 어순을 재배치한 유형 ② 및 동일한 줄표를 번역에 도입한 유형 ⑧에서는 40만-55만 어절 구간을, 기타 줄표를 없애고 어순을 그대로 둔 채 풀어쓴 유형 ③, 줄표가 있는 어구를 번역자가 누락한 유형 ④, 말줄임표로 변환한 유형 ⑤, 괄호로 처리한 유형 ⑥, 쉽표로 대체한 유형 ⑦에서는 25만-40만 어절 구간을 유의미한 코퍼스 크기의 결정 기준으로 확인할 수 있었다. 그러므로 우리는 최대 55만 어절이 확보되면 영어 원문의 줄표가 한국어 번역문에서 어떻게 번역될 것인가에 대한 답변이 예측 가능하다는 결론을 얻을 수 있다.

5. 맺음말

코퍼스를 활용한 통계 분석은 생성된 자료가 보여주는 특징적 경향을 있는 그대로 알려준다는 장점이 있다. 그래서 우리는 통계 처리에 의지해서 대규모 자료가 ‘말하고자 하는 어떤 경향’을 알아차리게 된다. 그러나 그것은 결과로서의 진술일 따름이며, 과정에 대한 해명이 되지 못한다. 이러한 맥락에서 우리는 영어 원문과 한국어 번역문으로 구성된 25만 어절 규모의 병렬코퍼스를 활용하여 영-한 번역문의 특징이 어떠한 번역 과정에서 비롯된 것인가를 분석한 바 있었다(김정우 2011). 그런데 영어 원문과 한국어 번역문으로 이루어진 병렬코퍼스의 정성적 분석에서 도출한 논의가 신뢰할 만한 결과가 되려면, 과연 코퍼스의 규모를 어느 정도까지 늘려야 할까? 통상적인 연구 환경에서 자료의 규모를 무한정 늘리기란 어려우므로, 분석 결과의 일반성을 담보할 수 있는 코퍼스의 적정 규모를 제안할 수 있다면 이 방면의 연구에 적잖게 기여할 수 있을 것으로 생각한다. 지금까지 논의된 내용을 요약하면서 결론에 갈음하기로 한다.

우리는 먼저 25만 어절의 병렬코퍼스를 대상으로 한 정성적 분석 결과(김정우 2011)를 요약하여 제시한 다음, 양적 인자의 증가에 따라 정성적 분석의 결과가 어떻게 변화하는가를 관찰하기 위해, 15만 어절씩 자료를 추가하여 누적 100만 어절에 이를 때까지 모두 다섯 단계에 걸쳐 분석 결과의 변화 추이를 비교하면서, 양적인 인자의 누적된 변화가 분석 결과에 질적으로 의미 있는 영향을 미치는, 곧 분석 결과의 일반성을 담보할 수 있는 적정 코퍼스의 규모를 확인해 보았다.

이상과 같은 조사와 분석의 결과로 우리는 다음과 같은 결론을 내릴 수 있었다.

첫째, 코퍼스 크기의 확대에 따라 분석 결과가 크고 작은 변화를 보였지만 큰 틀에서 볼 때 25만 어절의 코퍼스에 대한 조사 결과에서 추출한 번역 보편소와 관련된 내용, 곧 ‘때문’과 선택적 명시화, 장형사동 구문과 구조적 명시화, 문장 접속부사와 선택적 명시화 및 줄표와 규범화의 존재를 부인할 만한 것은 아니었다.

(1) 우리는 25만 어절에 대한 조사를 통해 의존명사 ‘때문’이 (원문에 대응하는 어구가 없는데도 문맥에서 유추한) 영형태를 번역한 경우가 약 45%에 달한다는 사실에서 선택적 명시화의 번역 보편소를 확인했는데, 유의미한 코퍼스 크기의 결정 기준이 되는 40만 어절에 대한 조사에서도 약 35%에 달하는 상대 빈도를 보였다.

(2) 우리는 25만 어절에 대한 조사를 통해 장형사동 구문의 일부(6.0%)가 단형사동 구문으로 교체될 수 있다는 사실과, 일반동사를 장형사동 구문으로 번역한 경우가 50% 가까이 된다는 사실에서 구조적 명시화의 번역 보편소를 확인했으며, 번역자가 임의로 도입한 장형사동 구문이 상당 분량(36%)에 달한다는 사실에서 장형사동 구문의 용법 확장을 확인했다. 이러한 제반 조사 결과는 유의미한 코퍼스 크기의 결정 기준이 되는 70만 어절과 40만 어절에 대한 조사에서도 각각 6.3%(교체 빈도), 약 50%(일반 동사 번역) 및 약 33%(의역) 등으로 거의 비슷하게 나타났다.

(3) 우리는 25만 어절에 대한 조사를 통해 문장접속부사 ‘그러나’, ‘하지만’ 및 ‘그런데’가 문맥에서 유추한 영형태에서 번역된 경우가 적잖게 나타난다는 사실(약 33%, 21%, 46%)에서 선택적 명시화의 번역 보편소를 확인했다. 이러

한 조사 결과는 유의미한 코퍼스 크기의 결정 기준이 되는 40만 어절과 55만 어절에 대한 조사에서도 각각 약 22%, 27% 및 53% 등으로 나타났다.

(4) 우리는 25만 어절에 대한 조사를 통해 영어 원문의 줄표가 한국어 번역 문에서 대략 일곱 가지 유형으로 나타나면서도 유독 줄표를 그대로 가져오는 번역 용례가 없었다는 점에 주목하여, 규범화의 번역 보편소를 확인했다. 이러한 결과는 코퍼스의 크기를 100만 어절까지 확대한 조사에서도 거의 그대로 유지되었다.²⁹⁾

둘째, 그러나 세부적인 사항, 예컨대 번역 연원의 상대 순위 등에 대해서는 수정된 진술이 필요하다. 25만 어절에 대한 조사 결과와 유의미한 규모의 (누적) 코퍼스에 대한 조사 결과를 비교해 기술해보자.

(1) 의존명사 ‘때문’의 번역 연원 조사에서, 가장 큰 차이는 25만 어절일 때 영형태의 번역 빈도가 가장 높았는데 유의미한 코퍼스 규모인 70만 어절일 때는 접속사의 번역 빈도가 가장 높게 나타났다는 점이다.

(2) 장형사동 구문의 경우는 코퍼스 크기의 확대에도 불구하고 전반적으로 유사한 결과를 얻었지만, 유의미한 코퍼스 규모인 70만 어절일 때는 사역동사 구문의 번역이 조금 증가하고 의역에서 비롯된 번역이 조금 감소한 사실을 알 수 있다(각각 약 3%).

(3) 문장접속부사 ‘그러나’의 번역 연원 조사에서, 눈에 띄이는 차이는 40만 어절일 때 접속사 ‘but’의 번역이 약 20% 정도 증가하고 영형태와 접속사 ‘however’의 번역이 각각 10% 정도씩 감소했다는 점이다. ‘하지만’의 번역 연원 조사에서, 가장 큰 차이는 55만 어절일 때 영형태의 번역이 약 6% 정도 증가했다는 점이다. ‘그런데’의 번역 연원 조사에서, 눈에 띄이는 차이는 55만 어절 때 영형태가 6% 증가하고 ‘but’이 4% 정도 감소했다는 점이다.

(4) 영어 원문 줄표의 한국어 번역 양상에 대한 조사에서, 눈에 띄이는 차이는 55만 어절일 때 새로운 문장으로 다시 쓴 경우와 어순을 유지한 경우 및 괄호로 변환한 경우가 약 2~3% 정도씩 증가한 반면, 줄표를 생략하거나 문장 전체를 누락시킨 경우와 말줄임표로 변환한 경우 및 쉼표로 변환한 경우가 약

29) 예외는 전체 3,940건 중의 8건으로 기존 논의에 거의 영향을 미치지 못하는 수치이다.

3~5% 정도씩 감소했다는 점이다.

셋째, 이론(가설)의 일반성을 담보할 수 있는 코퍼스의 적정 규모는 조사 항목에 따라 최소 40만 어절부터 최대 70만 어절로 나타났다. 그러므로 적어도 우리가 조사한 4개 항목에 관한 한, 70만 어절을 넘는 코퍼스의 경우는 100만 어절까지 분석 결과에 차이가 없는 것으로 판단할 수 있다.³⁰⁾

참고문헌

- 곽성희 (2000) 「텍스트성과 번역 전환 - 영한 번역을 중심으로」, 『번역학연구』 1(1): 93-118.
- 김도훈 (2010) 「대시(dash)와 줄표의 비교 및 영한 번역 전략」, 『번역학연구』 11(3): 41-64.
- 김순영 (2005) 「Evidence of Explicitation in Texts Translated from English into Korean: A Corpus-based Pilot Study」, 『번역학연구』 6(1): 143-166.
- 김영신 (2009) 「영한 번역에서 결속장치 전환 연구 - 접속사를 중심으로」, 『통번역교육연구』 7(1): 97-113.
- 김윤경 편 (1996) 『응용언어학과 통계학』, 서울: 한국문화사.
- 김정우 (2009) 「한국어 번역문의 구조적 명시화 - 특정 보어 구문의 열세적 분포와 관련하여 -」, 『번역학연구』 10(3): 77-97.
- _____ (2011) 「한국어 번역문의 중간언어적 특성」, 『번역학연구』 12(1): 75-122.

30) 코퍼스 크기의 증가에 따라 정성적 분석 결과가 달라진 것은 여러 가지 요인, 예컨대 원문 텍스트 유형(문학 텍스트와 비문학 텍스트)이나 원저자의 문체, 번역자의 문체 등이 작용한 결과일 것이다. 텍스트의 유형을 한정시킨다든가 특정 번역자의 번역으로 자료의 범위를 한정한다면, 분석 결과의 일반성 확보에 기여하는 좀더 정밀한 유의미 코퍼스의 규모 산출이 가능할 것이다. 다만, 25만 어절 코퍼스 분석 때는 문학 대 비문학 텍스트의 비율을 대략 1:1로 맞추었으나 추가된 코퍼스 분석 때는 이 비율이 1:2가 되어 비문학 텍스트의 비중이 높았음을 밝혀둔다. 문학과 비문학 텍스트의 구성이 달랐다면 코퍼스의 크기에 관한 결론도 달라졌을 수 있음을 인정한다. (익명의 심사자 지적)

- 김혜영 (2009) 「국어 번역 글쓰기의 연구」, 고려대 국문과 박사학위 논문.
- 김효중 (1998) 『번역학』, 서울: 민음사.
- 남원준 (2012) 「코퍼스를 활용한 번역연구에 관한 담론 - 선행연구를 중심으로」, 『통번역교육연구』 10(1): 147-172.
- 류현주 (2009) 「번역투와 번역자투」, 『번역학연구』 10(2): 7-22.
- 안동환 옮김 (2008), 『코퍼스기반 번역학』, 서울: 동인. [Sara Laviosa (2002) *Corpus-Based Translation Studies: Theory, Findings, Applications*, Amsterdam: Radopi]
- 오미영 옮김 (2004) 『대조언어학』, 서울: 제이앤씨. [石綿敏雄·高田誠 (1990) 對照言語學, おうふう社: 東京]
- 이근희 (2005) 「영한 번역에서의 ‘번역투’ 연구」, 세종대 영문과 박사학위 논문.
- 이영옥 (2003) 「한영 간 장소이동 구문 번역의 문제」, 『번역학연구』 4(1): 51-71.
- 이영훈 외 (2010) 「비교코퍼스에 기반한 불한번역 문학텍스트의 언어적 특성 연구」, 『번역학연구』 11(1): 161-190.
- 이창수 (2011) 「엔그램 분석을 통한 번역한국어와 비번역한국어 간의 어휘묵음 유형 차이 사례연구」, 『통번역학연구』 15권. 317-340.
- 전성기 (1996) 『불한 번역 대조 분석』, 서울: 어문학사.
- 정연일·남원준 옮김 (2005) 『번역학 입문-이론과 적용』, 서울: 한국외국어대학 교출판부.(*Introducing Translation Studies*).
- 정호정 (2003a) 「코퍼스 중심의 번역학 연구」, 『번역학연구』 4(2): 71-88.
- _____ (2007) 『제대로 된 통역·번역의 이해』, 서울: 한국문화사.
- 조의연 (2009) 「병렬 말뭉치에 기반한 번역학 연구」, 『번역학연구』 10(2): 207-246.
- _____ (2012) 「번역가의 문체선택 연구」, 『번역학연구』 13(5): 197-221.
- 조재범 (2008) 「코퍼스를 통한 After all의 용례 분석 및 번역 교육에의 적용」, 『통번역교육연구』 6(1): 5-17.
- 조준형 (2012) 「병렬코퍼스에서의 맥락 탐색의 의미와 한계」, 『번역학연구』 13(5): 223-246.
- 최승권·김영길 (2010) 「번역 말뭉치로부터 추출한 어휘 번역 패턴의 의미 분

류와 자동번역시스템에의 활용」, 『번역학연구』 11(3): 277-301.

최진실 (2010) 「접속 부가어의 영한 번역 연구」, 부산대 영문과 박사학위 논문.

_____ (2011) 「번역의 방향과 접속 부가어의 의미」, 『번역학연구』 12(2): 207-232.

Baker M. (1993) 'Corpus Linguistics and Translation Studies: Implications and Applications', in Baker, G. Francis and E. Tognini-Bonelli (eds), 233-250.

_____ (1998) (ed) *Routledge Encyclopedia of Translation Studies*, London: Routledge.

Jacobson, R (1959) 'On Linguistic Aspects of Translation', in R. A. Brower, (ed) *On Translation*, Massachusettes: Harvard UP.

Nida, E. A. & C. R. Taber (1969) *The Theory and Practice of Translation*, Leiden, Brill.

Pym, Anthony (1998) *Method in Translation History*, Manchester: St. Jerome, UK.

Toury, G. (1995) *Descriptive Translation Studies and Beyond*, Amsterdam and Philadelphia: John Benjamins.

부록

(1) 구간별 변화량 일람표

1. ‘때문’의 번역 연원³¹⁾

유형/증감	25만-40만	40만-55만	55만-70만	70만-85만	85만-100만
a	-9.6	-2.1	-2.0	-0.5	+0.1
b	+2.7	+1.1	+0.4	-0.8	-0.7
c	+1.2	-1.0	+2.2	+1.0	+0.4
d	+6.7	+2.2	-0.4	+0.2	-0.1
e	-0.8	-0.1	-0.1	+0.1	+0.3
f	-0.2	-0.1	-0.1	0.0	0.0

2. 단형사동 교체 가능형

유형/증감	25만-40만	40만-55만	55만-70만	70만-85만	85만-100만
교체 가능	-0.6	+0.4	+0.5	+0.3	0.0

3. 장형사동의 번역 연원

유형/증감	25만-40만	40만-55만	55만-70만	70만-85만	85만-100만
a	+3.0	0.0	-0.4	+0.2	-0.2
b	0.0	-1.5	+1.8	+1.0	-0.3
c	-3.0	+1.5	-1.4	-1.2	+0.5

4. 문장접속부사 직접 대조 대 간접 대조의 비율

유형/증감	25만-40만	40만-55만	55만-70만	70만-85만	85만-100만
간접/직접	-5.0	+3.3	+0.1	-1.2	+0.8

31) 아래 도표의 유형 표시에 보이는 알파벳 기호는 표기의 편의상 사용한 것이다. 이들 기호가 지칭하는 구체적인 내용은 본문의 도표를 참고하면 된다.

5. 문장접속부사 ‘그러나’의 번역 연원

유형/증감	25만-40만	40만-55만	55만-70만	70만-85만	85만-100만
a	+20.9	-1.5	-1.9	-2.1	-1.1
b	-11.8	+4.2	-0.4	-0.5	+0.9
c	-10.0	-2.2	+1.9	+2.1	-0.7

6. 문장접속부사 ‘하지만’의 번역 연원

유형/증감	25만-40만	40만-55만	55만-70만	70만-85만	85만-100만
a	+0.6	-3.1	-0.7	-2.0	-1.3
b	+1.3	+5.0	+0.5	+2.6	-0.1
c	-0.1	-0.9	-0.1	0.0	-0.1
d	+0.2	+1.0	-0.2	+0.2	-0.7

7. 문장접속부사 ‘그런데’의 번역 연원

유형/증감	25만-40만	40만-55만	55만-70만	70만-85만	85만-100만
a	-3.5	+10.4	-1.4	+0.1	+4.4
b	+2.2	-6.5	-1.2	-0.8	-1.0
c	+2.9	-1.2	-2.0	-0.6	-0.8

8. 줄표(dash)의 번역 연원

유형/증감	25만-40만	40만-55만	55만-70만	70만-85만	85만-100만
a	-1.4	+4.1	-2.0	-1.1	0.0
b	+0.7	-1.0	+0.8	0.0	0.0
c	+3.0	-0.8	+2.3	0.0	+0.1
d	-2.5	0.0	-0.9	0.0	0.0
e	-3.6	-0.9	+1.5	0.0	-0.3
f	+4.5	-1.6	-0.5	0.0	+0.2
g	-1.7	0.0	-0.1	0.0	0.0
h	0.0	+0.2	0.0	0.0	0.0

(2) 항목별 기술통계량 일람표

1. ‘때문’의 번역 연원

기술통계량

	N	최소값	최대값	평균	표준편차
a2	5	-9.60	.10	-2.7300	3.97863
b2	5	-.80	2.70	.5400	1.44326
c2	5	-1.00	2.20	.7600	1.17813
d2	5	-.40	6.70	1.7200	2.96429
e2	5	-.80	.30	-.1200	.41473
f2	5	-.20	.00	-.0800	.08367
유효수(목록별)	5				

2. 단형사동 교체 가능형

기술통계량

	N	최소값	최대값	평균	표준편차
a3	5	-.60	.50	.1200	.44386
유효수(목록별)	5				

3. 장형사동 번역 연원

기술통계량

	N	최소값	최대값	평균	표준편차
a4	5	-.40	3.00	.5200	1.40428
b4	5	-1.50	1.80	.2000	1.26293
c4	5	-3.00	1.50	-.7200	1.75414
유효수(목록별)	5				

4. 문장접속부사 직접 대조와 간접 대조 비율

기술통계량

	N	최소값	최대값	평균	표준편차
a5	5	-5.00	3.30	-.4000	3.04877
유효수(목록별)	5				

5. 문장접속부사 ‘그러나’의 번역 연원

기술통계량

	N	최소값	최대값	평균	표준편차
a6	5	-2.10	20.90	2.8600	10.09198
b6	5	-11.80	4.20	-1.5200	6.05285
c6	5	-10.00	2.10	-1.7800	4.93731
유효수(목록별)	5				

6. 문장접속부사 ‘하지만’의 번역 연원

기술통계량

	N	최소값	최대값	평균	표준편차
a7	5	-3.10	.60	-1.3000	1.33744
b7	5	-.10	5.00	1.8600	2.02559
c7	5	-.90	.00	-.2400	.37148
d7	5	-.70	1.00	.1000	.62450
유효수(목록별)	5				

7. 문장접속부사 ‘그런데’의 번역 연원

기술통계량

	N	최소값	최대값	평균	표준편차
a8	5	-3.50	10.40	2.0000	5.51679
b8	5	-6.50	2.20	-1.4600	3.14293
c8	5	-1.20	2.90	.4600	1.85638
유효수(목록별)	5				

8. 줄표(dash)의 번역 연원

기술통계량

	N	최소값	최대값	평균	표준편차
a9	5	-2.00	4.10	-.0600	2.44683
b9	5	-1.00	.80	.1000	.72111
c9	5	-.80	3.00	.9200	1.63615
d9	5	-2.50	.00	-.6300	1.08950
e9	5	-3.60	1.50	-.6300	1.86628
f9	5	-1.60	4.50	.5200	2.33174
g9	5	-1.70	.00	-.3600	.75033
h9	5	.00	.20	.0400	.06944
유효수(목록별)	5				

[Abstract]

How Much Do Quantitative Factors Affect Qualitative Analyses in Corpus-based Translation Studies?

Kim, Jeong-Woo
(Kyungnam University)

This paper aims at elucidating what size of corpus can produce the reliable qualitative analyses when the parallel corpus, composed of English original and Korean translated texts, is used. To reach the goal, we have divided the size of corpus into 5 levels from a quarter-million to one million (phonological) words. At each level, the number of words has been increased by one hundred fifty thousand words, i.e., 250,000, 400,000, 550,000, 700,000, 850,000, and 1,000,000 words. Then, we have examined the major differences between the levels. The results obtained from our investigation are as follows:

First, with reference to the translation source of the Korean bound noun *ttaemun*(reason or ground), the zero-morph translation is most frequent in a quarter-million corpus level, while the frequency of the conjunctive translation is the highest in the seven hundred thousand corpus level. This indicates that at least, the corpus size of seven hundred thousand words is necessary to get a meaningful analysis of the bound noun *ttaemun*. Second, although the differences between the five levels are not significant, the translation of the long-form causative construction becomes more frequent in the seven hundred thousand corpus level while the frequency of the text-free translation decreases more or less. Third, in the case of the translation source of the Korean conjunctive *geureona*(but), the translation frequency of conjunctive ‘but’ increases by 20 percent in the four hundred thousand corpus level while the translation of either zero morph or conjunctive ‘however’ decreases by 10

percent in the same corpus level. On the other hand, in the case of the Korean conjunctive *hajiman*(yet or but), certain significant change of translation frequency occurs in the five hundred fifty thousand corpus level. Finally, concerning the translation of the English dash mark ‘-’ into Korean, the five hundred fifty corpus level shows a significant result. For example, the dash mark disappears in many Korean texts, or the contents after the dash mark is rewritten as a new Korean sentence.

In conclusion, the reasonable size of corpus, which can be developed into a hypothesis or theory, can vary from four hundred thousand words minimally to seven hundred thousand words maximally according to our investigation. Furthermore, the corpus size over seven hundred thousand words does not make any difference on the qualitative analyses of the 4 items thoroughly investigated in this paper.

▶ Key Words: translation universals, simplification, explicitation, normalization, parallel corpus, Korean translated texts from English, quantitative factors, qualitative results

김정우
경남대학교 국어국문학과 교수
kjwn@kyungnam.ac.kr
관심 분야: 번역이론, 번역사

논문투고일: 2013년 1월 31일

심사완료일: 2013년 3월 9일

게재확정일: 2013년 3월 13일