

ParaConc를 이용한 영·한 병렬말뭉치 분석 연구

박 명 수
(상명대)

1. 서론

1960년대 초반 최초의 코퍼스라 불리는 Brown 말뭉치(Brown Corpus)가 세상에 선을 보인 지 60년 가까이 되었다. 이후로 컴퓨터 기술의 발전에 힘입어 수 없이 많은 학자들이 다양한 말뭉치를 이용해 언어 관련 연구를 수행하고 있다. 원어민의 직관에 의존하기 보다는 신뢰할 수 있는 방대한 양의 언어 데이터 수집과 활용이 가능해지면서 언어 관련 연구는 큰 변화를 겪어왔다. 언어학, 외국어교육, 번역학 등에서 말뭉치에 대해 점점 관심을 많이 갖는 이유는 많은 양의 데이터를 수집한 후 코퍼스 데이터로 구축하면 생생한 언어 데이터를 기반으로 다양한 분석이 가능하고, 이전까지는 할 수 없었던 거대한 규모의 데이터를 분석할 수 있기 때문이며(박명수 2017), 번역 연구에서 뿐만 아니라, 번역 교육 및 자동번역과 같은 다양한 연관 분야에서 응용될 수 있는 유용한 도구로 인식되고 있다(조준형 2012).

100만 단어로 시작한 Brown 말뭉치의 등장 이후, 말뭉치는 규모 면에서 눈

부신 성장을 했다. 오늘날 영어권에서 가장 활용도가 높은 코퍼스 중 하나인 The British National Corpus (BNC)는 토큰(token)이 1억이며, The Corpus of Contemporary American English (COCA)는 5억 6천이 넘는다. 규모뿐 아니라, 말뭉치의 종류도 원어민 데이터에서 벗어나 외국어를 공부하는 학습자들이 쓴 글이나 말을 기반으로 구축한 학습자 코퍼스(learner corpus), 특정 장르나 도메인 등에 국한해 분석을 하고자 텍스트를 모아 구축한 특수 코퍼스(specialized corpus)등도 다양한 목적으로 활용되고 있다. 이러한 단일어 기반 코퍼스(monolingual corpus) 뿐만 아니라, 두 개 이상의 언어로 된 텍스트를 수집해서 구축한 다언어 코퍼스(multilingual corpus)도 등장했다.

본 연구의 주제인 병렬말뭉치(parallel corpus)는 이러한 다언어 코퍼스에 해당하는데, 원문 텍스트와 원문 내용을 정확하게 옮긴 번역 텍스트가 나란히 병렬로 구축되어 있는 경우를 지칭한다. 위에서 열거한 다양한 종류의 말뭉치들도 번역 연구에 활용될 수 있지만, 실제 번역학에서 가장 관심을 두어야 하는 말뭉치가 바로 병렬말뭉치이다.

병렬말뭉치에 대한 연구는 기계번역(machine translation), 자연어 처리(natural language processing) 등의 영역과도 아주 밀접한 직, 간접적 연관성이 있어 이에 대한 연구의 활성화가 절대적으로 필요하다. 병렬 말뭉치는 원문과 번역문이 동시에 말뭉치로 구축되어야 하기에, 번역이 이미 이루어진 텍스트가 존재해야 가능하다. 우리나라에서는 문학작품과 신문기사의 원문과 번역문을 병렬 말뭉치로 구축한 사례가 많은데, 본 연구에서도 영자신문의 사설 원문과 번역문을 대상으로 영한 병렬말뭉치를 구축했다. 기존 병렬 말뭉치 연구들이 말뭉치 구축과 구축 과정에서 발생하는 정렬(alignment) 문제에 대한 해결책 제시를 다루거나, 세종말뭉치에서 구축한 한영 병렬말뭉치 이외에는 병렬말뭉치 분석에서 말뭉치 분석도구(concordancer)를 사용하지 않은 채 연구자가 거의 수동적인 분석을 한 경우가 많았다. 본 연구는 병렬말뭉치 분석도구로 널리 알려진 'ParaConc'를 이용해 영한 병렬 말뭉치 분석을 시행한다는 점에서 기존 연구와 차별점을 찾을 수 있다. ParaConc를 이용한 영한 병렬말뭉치 분석을 통해, 서양어 중심으로 개발된 ParaConc의 활용 가능성을 분석해본다.

2. 선행연구 분석

개인용 컴퓨터의 대중화에 힘입어, 코퍼스 언어학(Corpus Linguistics)을 기반으로 한 연구의 가능성이 열린 셈이니, 그 역사는 그리 길지 못하다. 번역에서 말뭉치에 기반한 연구의 가능성이 언급된 시점은 대체로 1990년 초반부터이다. 베이커(Baker 1993)가 번역 텍스트 연구를 위한 코퍼스 활용법을 처음 논했고, 영국 맨체스터 대학교의 통번역학 센터에서 번역영어코퍼스(Translational English Corpus) 구축을 위해 수행한 프로젝트에서 코퍼스 언어학의 다양한 분석 기법이 번역학에 도입되었다(Stubbs 1996). 국내에서도 말뭉치 기반 번역 연구도 2000년대에 들어서면서 활성화되고 있는데, 이 가운데 병렬 말뭉치를 기반으로 한 연구는 다른 말뭉치 기반 연구에 비하면 그 수가 적어 아직까지 발전의 여지가 많다고 할 수 있다. 먼저 우리나라에서 구축된 영-한/한-영 병렬말뭉치 현황을 살펴보고, 이를 기반으로 한 연구를 분석해본다.

2.1 국내 영-한/한-영 병렬말뭉치 구축 현황

영어를 원문으로, 한글을 번역문으로 또는 그 반대의 경우로 구축된 병렬말뭉치는 개별 연구자들이 직접 구축한 경우와 국가나 기관에서 추진하는 프로젝트로 여러 명의 연구자들이 구축한 경우로 나눌 수 있는데, 개별 연구자가 구축한 병렬말뭉치는 해당 연구 논문이외에는 공개되는 경우가 드물다. 아래 <표 1>에 나타난 바와 같이 세종 한영 병렬말뭉치와 KAIST에서 구축한 병렬말뭉치가 우리나라를 대표할 수 있는 병렬말뭉치라 할 수 있다.

세종 한·영 병렬말뭉치는 1998년부터 2007년까지 국어 정보화 중장기 발전 계획에 따라 진행된 ‘21세기 세종계획’이 10년의 성과물 중 하나이다. 원래 ‘세종 말뭉치’는 ‘문어, 구어, 병렬, 역사’ 말뭉치로 구성되어 있는데, 이 중에서 본 연구의 주제인 병렬말뭉치의 경우 영어, 일본어, 러시아어, 프랑스어, 중국어와 각각 병렬 구조로 만들어 공개했다(유현경·황은하 2009).

〈표 1〉 한·영 병렬말뭉치

말뭉치 이름	규모	공개연도	분석기
세종 한영 병렬말뭉치 (원시)	약 500만 어절	2007	hepman
세종 한영 병렬말뭉치 (형태소 분석)	약 100만 어절	2007	hepman
KAIST 영한 다국어 코퍼스	6만 문장쌍	2005	없음

세종 한영 병렬말뭉치는 ‘21세기 세종계획 특수말뭉치 소분과 병렬말뭉치 과제’ 중에서 가장 규모가 큰 병렬말뭉치이며 현재까지 우리나라에서 가장 잘 알려진 대규모 병렬말뭉치이다. 1998년부터 3단계에 걸쳐 진행되었는데 원시 병렬말뭉치의 경우 잡지, 책, 정부 문서 등과 같은 출판물, 연설문 등과 같은 비출판물을 수집해 4,763,932 어절 규모로 구성되어 있다(서상규 2008) 세종 한영 병렬말뭉치는 다른 분석도구로는 분석이 안되고 ‘한·영 병렬말뭉치 용례 검색기’인 ‘hepman’을 사용해야 한국어와 영어의 대조 분석 등 연구에 활용할 수 있다. KAIST 영한 다국어코퍼스는 Semantic Web Research Center(SWRC)에서 구축한 것으로 영어와 한국어를 문장 단위로 정렬한 6만 문장쌍으로 구성되어 있는데 개별적인 문장의 모음으로 아무런 전후맥락이 없고, 영어 원문에 대한 한글의 번역도 자연스럽게 못한 수준의 번역이어서 제대로 된 병렬말뭉치라 보기에 부족하다. 아울러 KAIST에서 동아일보의 사설을 원문으로 그리고 영어 번역문을 토대로 구축한 병렬코퍼스가 있는데, 이는 .rtf파일 형태로 되어 있지만 할 뿐 아무런 정렬이 되어 있지 않아, 실제 연구용으로 활용하려면 변환작업이 필요한 상태이다.

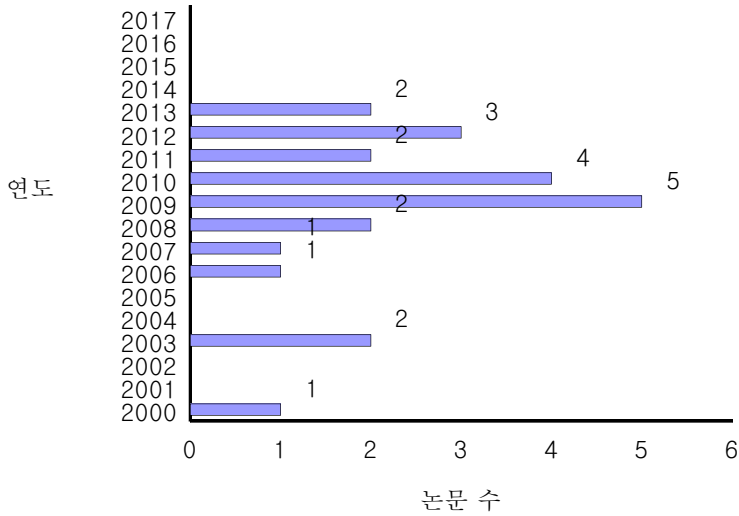
2.2 국내 병렬말뭉치 연구

우리나라 학술지에 게재된 영-한, 한-영 병렬말뭉치에 대한 연구를 검색하기 위해, 학술지 데이터베이스인 www.riss.kr과 www.dbpia.co.kr를 이용했다. 2017년 6월 말까지 “병렬코퍼스” “병렬말뭉치”로 검색한 결과 추출된 총 131편의 논문 중에서 영-한 또는 한-영 병렬말뭉치를 기반으로 한 연구만을 대상으로

- 1) KAIST에서 구축한 병렬말뭉치는 다음 웹페이지에서 무료로 다운로드가 가능하다.
http://semanticweb.kaist.ac.kr/home/index.php/KAIST_Corpus

아래와 같이 분석을 했다.

그림 1 한영/영한 병렬말뭉치기반 연구논문



위 그림 1에서 알 수 있듯이 우리나라 학술지에 실린 영-한, 한-영 병렬말뭉치에 대한 연구는 양적으로 볼 때 상당히 저조한 편이다. 2000년 이후로 2017년 6월 말까지를 기준으로 23편에 불과하다. 물론 앞서 언급했던 것처럼 분석 대상 연도에 학술지에 실린 말뭉치 기반 연구는 훨씬 많지만, 영어가 아닌 중국어, 일본어, 프랑스어, 독일어 등의 다른 외국어와 한글 병렬말뭉치 연구들이었고, 컴퓨터, 소프트웨어 공학 분야에서 진행한 병렬말뭉치 구축을 위한 시스템, 프로그램, 말뭉치 정렬(alignment)과 같은 기술적인 내용을 다룬 연구들이었다. 세종 한-영 병렬말뭉치가 공개된 시점이 2007년인데 2007년 이후로 2013년까지 매 해 한-영 병렬말뭉치 기반 연구들이 집중되어 있기는 하지만, 국가 차원에서 진행해서 공개한 한-영 병렬말뭉치를 기반으로 한 연구는 이 중에서 몇 개에 지나지 않는다.

23편 중에서 언어기반 연구가 아닌 경우를 제외하고, 주제별로 분류를 해보

면 아래 <표 2>와 같이 요약할 수 있다. 현재까지 영-한, 한-영 병렬말뭉치 기반 연구는 원문이나 번역문에 나타나는 특정 어구, 구문 등의 번역 사례를 분석하는 방식의 연구가 주류를 이루고 있다(송상헌 2012; 신혜정 2013; 유현경·안예리·양수향 2007; 이승아 2012; 이승아·배지연 2011; 정연창·김은일 2009; 조의연 2009). 번역 교육이나 한국어 또는 특정 외국어 교육을 위한 병렬말뭉치의 가치를 언급한 연구는 원론적인 수준에서 다루어졌다(남원준 2006; 신자영 2011). 그 이외 나머지 연구논문들은 병렬말뭉치의 구축과정을 설명한 기술적인 수준의 연구이거나 병렬말뭉치 활용의 의미를 설명한 연구이다. 병렬말뭉치 구축과 활용에 대한 연구가 2008년부터 2011년 사이에 집중된 이유는 특정 학술지의 특집호에서 세종계획 21의 성과의 의미를 다루었기 때문이다. 기술적이거나 이론적인 내용의 연구를 제외하고 실제 병렬말뭉치를 활용해 분석하고 결과를 도출한 연구는 이 중에서 11편에 불과하다.

<표 2> 한-영/ 영-한 병렬말뭉치 기반 연구 주제별 분류

주제		연구자	병렬말뭉치	텍스트
번역/ 언어 교육	번역교육	남원준(2006)	한-영/ 영-한	관광안내문
	중간언어 연구	신자영(2010)	한-영/ 영-한	소설
번역 기법	as if 구문	신혜정(2013)	영-한	소설
	be being + 서술형용사	이승아(2012)	영-한	소설
	관용구	김한샘(2009)	한-영	세종말뭉치 (연설문, 소설)
	by 수동문	이승아(2009)	영-한	뉴스
	능동-수동	정연창·김은일(2009)	한-영	소설
	조사 ‘가’ ‘는’	유현경·안예리·양수향(2007)	한-영	소설
	지시한정사	조의연(2009)	영-한	소설
	과거시제	송상헌(2012)	한-영	세종 말뭉치
은유 구문	이승아·배지연(2011)	영-한	뉴스	

번역이론	번역단위	주향아·조미희(2010)	한-영	뉴스
	번역단어 수	최정아(2003)	한-영/ 영-한	소설 20편
병렬말뭉치 구축 및 활용		민경모(2010)	한-외	특정되지 않음
		서정목(2010)	한-외	특정되지 않음
		박정열·차정원 (2008)	한-외	특정되지 않음
		서상규(2009)	한-외	특정되지 않음
		김홍규·정태구· 김정숙(2000)	한-외	특정되지 않음
		신자영·원미진 (2011)	한-외	특정되지 않음

위에서 언급한 11편의 연구에서 사용한 말뭉치는 소설이나 뉴스를 텍스트로 사용한 경우가 대부분이다. 말뭉치 연구에서 말뭉치의 규모를 정확하게 소개하는 것이 일반적인데 대부분의 연구들이 말뭉치의 규모에 대한 명확한 언급 없이 병렬말뭉치를 구축했고 연구 주제에 해당하는 문장을 추출해서 비교 분석했다. 말뭉치의 규모면에서 보면 최정아(2003)는 영-한 번역 소설 10편과 한-영 번역 소설 10편을 말뭉치로 구축해 연구에 사용했고, 이승아(2012)는 해리포터 소설 시리즈 총 7권을 병렬말뭉치로 구축해서 어느 정도 규모가 있지만, 이를 제외하고는 한 편의 소설을 말뭉치로 구축해 규모가 상당히 적다고 볼 수 있다. 이렇게 작은 규모의 병렬말뭉치 구축이 되풀이 되고 구축한 말뭉치가 공유될 수 없는 이유는 연구자 개인이 시도하기에는 비용과 시간 면에서 상당히 어려움이 따르고, 저작권 등의 문제로 인해 연구용 이외에 외부에 공개하는 것이 불가능하기 때문으로 풀이된다. 또 다른 이유로 한영 병렬말뭉치를 다루기 위한 마땅한 검색도구가 없어서 수작업으로 분석을 할 수 밖에 없기 때문에 대규모 말뭉치를 다루는데 한계가 있다(유현경·안예리·양수향, 2007). 세종 병렬말뭉치의 경우 ‘hepman’이라는 병렬말뭉치 분석도구도 함께 공개했지만, 현재까지 김한샘(2009)과 송상현(2012)만이 세종 한-영 병렬말뭉치로 연구를 수행했다.

현재까지 영-한, 한-영 병렬말뭉치에 대한 연구가 양적인 면에서 절대적으로 부족한 가장 큰 이유 중 하나는 바로 분석도구의 부재를 들 수 있다. 연구자가 많은 양의 병렬말뭉치를 컴퓨터가 읽어낼 수 있는 형식으로 구축하더라도

AntConc, MonoConc, WordSmith 등과 같은 단일어 코퍼스 분석도구로 다양한 말뭉치 분석이 가능하지 않다면 모두 수작업으로 할 수 밖에 없다. 이에 본 연구는 병렬말뭉치를 구축해서 이미 해외 많은 연구자들이 활용하고 있는 ParaConc를 사용해 영-한, 한-영 병렬말뭉치 분석의 가능성을 확인하고, 장·단점을 파악하고자 한다.

3. 연구 방법

3.1 병렬말뭉치 구축

본 연구에 사용한 병렬말뭉치는 우리나라의 대표 영자신문 중 하나인 코리아타임즈(The Korea Times)의 사설을 원문으로 한글 번역문과 함께 수집해서 구축했다. 코리아 타임즈 웹페이지에서 2014년 1월부터 2017년 4월까지 발행된 신문의 영문사설과 해당 사설의 한글 번역문으로 각각 293개 파일을 수집해 영-한 병렬말뭉치를 구축했다. 웹페이지에서 수집한 각각의 사설을 텍스트 파일로 변환, 저장한 후, 원문과 번역문을 문장 단위로 정렬하는 수작업을 거쳐 Parallel Corpus of English-Korean Editorials(PACOEK)를 구축하였다. 아래 <표 3>은 PACOEK에 대한 요약 정보이다.

<표 3> 병렬말뭉치 PACOEK 내용

말뭉치 구성	파일 수	규모	비고
영어 원문	293	150,071(토큰)	
한글 번역문	293	95,297(어절)	

3.2 병렬말뭉치 분석 도구

본 연구는 병렬말뭉치 분석도구로 ParaConc²⁾를 사용했다. 앞서 언급한 바

2) 본 연구에 사용한 ParaConc는 유료 버전으로, ParaConc에 대한 보다 자세한 정보는 www.paraconc.com이나 www.athel.com에서 확인할 수 있다.

와 같이 단일어 말뭉치 분석의 경우 코퍼스 연구에서 가장 많이 활용되고 있는 콘코댄서(concordancer)로 Word Smith와 AntConc가 있는데, 이는 문자 그대로 한 가지 언어만을 검색할 수 있다. 병렬말뭉치를 분석하려면 두 가지 이상의 언어 텍스트를 동시에 검색, 분석할 수 있어야 한다. 이를 위해서 ParaConc를 사용했다. ParaConc는 대조분석, 번역과정 자체를 연구하기 위해 번역문에 대한 연구를 수행하는 언어학자 및 연구자들을 위한 분석도구이다(Barlow, 2001).

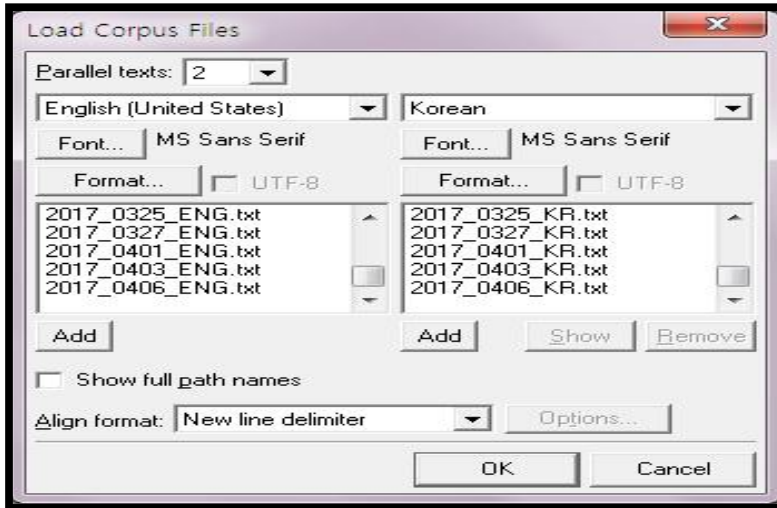
4. ParaConc를 이용한 병렬말뭉치 분석

본 장에서는 ParaConc를 사용해서 병렬말뭉치를 번역교육에 활용하는 가능성에 대해 살펴본다. 기능적인 측면으로 볼 때 ParaConc는 단일 말뭉치 분석도구들보다 제한적이다. 빈도수(frequency)와 연어(collocation) 분석을 제외하고 n-gram 분석이나 keyword 분석기능은 전혀 없다. 하지만 병렬말뭉치를 사용한다는 것은 두 언어를 상호 비교하는 것이고, 특히 특정 어휘나 어구가 어떻게 표현되었는지를 비교할 수 있기 때문에 단일 말뭉치 분석도구 사용과는 목적 자체가 다르다는 점을 인식해야 한다. 따라서 본 장에서는 ParaConc의 기능 중에서 ‘Translation’ ‘Hot Words’ ‘Parallel Search’ ‘Advanced Search’를 이용한 분석한 결과를 제시한다.

4.1 병렬말뭉치 탑재

텍스트 파일로 변환한 영어 원문과 한글 번역문 각각 293개 파일을 ParaConc에 그림 2와 같이 탑재한다. 탑재할 때 화면에서처럼 언어를 좌우에 각각 영어, 한글로 지정한 후 해당하는 언어 파일을 탑재해야 한다.

그림 2 ParaConc에 병렬말뭉치 탑재



탑재를 하고 나면, 화면 맨 하단에 탑재된 파일 수와 말뭉치를 구성하는 토큰 수를 확인할 수 있다.

4.2 일반 검색 및 고급 검색 기능

말뭉치 분석에서 가장 먼저 수행하는 것이 어휘의 빈도수 분석이다. 아래 그림 3은 ParaConc를 사용해 원문과 번역문에 등장하는 어휘를 빈도 순서로 추출한 결과화면이다. 영어 텍스트에서 상위를 차지하는 어휘는 대부분 기능어(function word)이고, 한글 텍스트에서도 내용어(content word)보다는 조사나 어미 등의 기능어들이 상위를 차지하고 있다. 영어는 모든 단어를 띄어 쓰지만 한글은 어미, 조사 등이 단어와 함께 쓰이는데다가 영어와 달리 독특한 띄어쓰기 방식으로 ParaConc의 어휘 빈도 분석 결과가 생각한 것과 다르게 보인다.

그림 3 ParaConc 어휘 빈도 분석 결과

English (United States)			Korean		
Count	Pct	Word	Count	Pct	Word
9848	6.8610%	the	902	0.9420%	이
4379	3.0508%	to	767	0.8010%	있다
4008	2.7924%	of	608	0.6360%	수
3220	2.2434%	and	534	0.5577%	것이다
2804	1.9535%	a	497	0.5190%	한다
2571	1.7912%	in	438	0.4574%	것은
1862	1.2972%	is	427	0.4459%	있는
1631	1.1363%	for	424	0.4428%	박
1562	1.0862%	that	392	0.4094%	더
1237	0.8618%	it	390	0.4073%	대한
1104	0.7692%	as	340	0.3551%	그러나
998	0.6953%	on	316	0.3300%	한
937	0.6528%	with	295	0.3081%	그
897	0.6346%	...	290	0.3026%	하고

이러한 영어와 한글의 차이는 아래 일반 검색을 해보면 잘 알 수 있다. 먼저 일반 검색으로 ‘정부’라는 단어를 일반 검색하면 단 70개 밖에 추출되지 않는다. 이는 우리 한글에서 ‘정부’라는 명사는 ‘정부의’ ‘정부가’ ‘정부를’ ‘정부에서’ 등과 같이 다양한 어미가 붙어 의미와 기능이 달라지지만, 영어의 경우에는 단어의 위치가 의미와 기능을 결정하기 때문이다. 그래서 한글을 검색하기 위해서 그림 4와 같이 고급검색(advanced search)의 ‘구문검색(search syntax)’중에서 ‘정규식 표현(regular expression)’ 검색 기능을 사용해야 ‘정부’의 정확한 빈도를 확인할 수 있다. 이는 한국어와 일본어의 경우 단어의 경계가 영어와 다른 점으로 인해 발생하는 문제이다. 또한 말뭉치 구축에서 한국어, 일본어와 같은 아시아 언어의 문자 인코딩 방식이 표준화 되지 않은 경우에 분석에 문제가 생길 가능성이 높다.³⁾

3) 한국어와 같이 아시아어가 포함된 다국어 코퍼스를 구축하는 경우에 문자 인코딩 문제 해결을 위해 일반적으로 유니코드(Unicode)방식이 권유된다(McEnery & Xiao 2004).

그림 4 ParaConc 고급검색의 정규식 표현 검색 설정화면

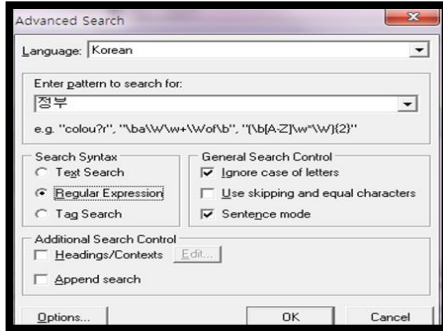
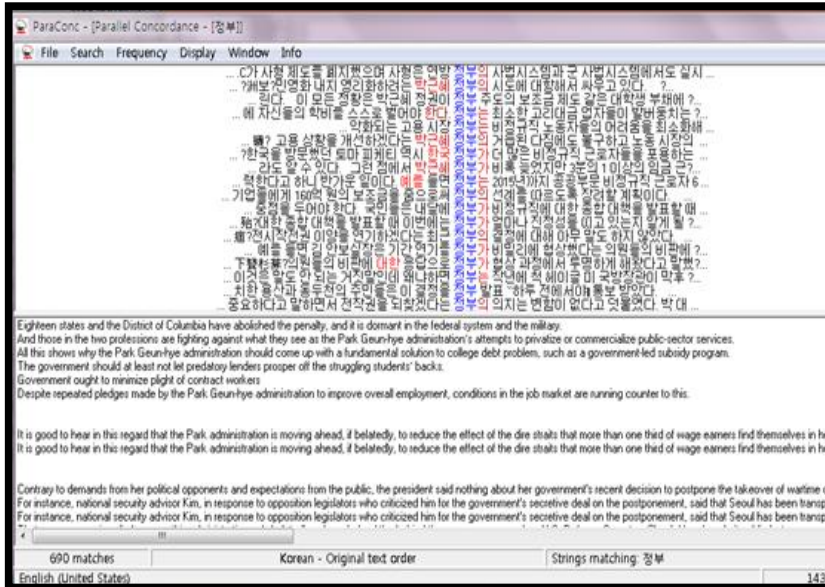


그림 4에서 설정한 정규식 표현(regular expression)으로 ‘정부’를 검색한 결과는 그림 5에서 확인할 수 있다. 앞서 ‘정부’를 일반 검색하면 결과가 70개 밖에 되지 않는다고 했는데, ‘정규식 표현’에 따른 ‘정부’ 검색의 결과는 690개나 된다. ParaConc로 우리 한글을 검색할 때는 ‘정규식 표현(regular expression)’을 이용해서 정확하게 의도한 내용을 검색할 수 있다는 것을 알 수 있다.

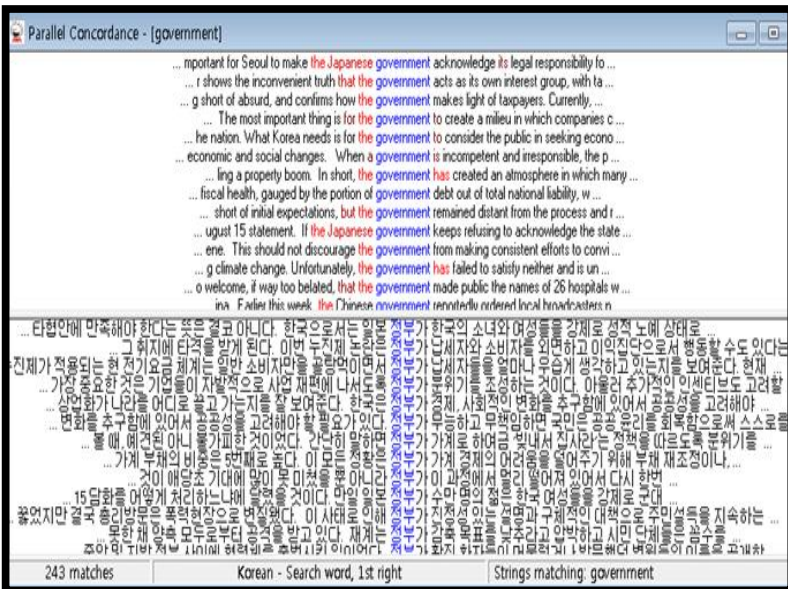
그림 5 ParaConc 정규식 표현 검색 결과화면



ParaConc 또는 Multiconcord (Wools 1997) 등 다른 병렬말뭉치 분석도구들이 알파벳으로 된 유럽 언어 분석을 전제로 개발되었기 때문에, 근본적으로 문자 언어가 알파벳이 아닌 한국어, 일본어, 중국어 등의 언어를 완벽하게 반영하는데 무리가 있을 수밖에 없다. 우리나라보다 병렬말뭉치 기반의 연구가 활성화된 중국어의 경우 문자 언어 및 문법 등의 근본적인 차이로 인한 병렬말뭉치 정렬 및 분석도구 개발을 위해 다양한 시도를 통해 문제를 해결해오고 있다. 예를 들어 릭순(Lixun 2001)은 2000년에 3년간의 노력 끝에 ‘E-C Concord’라는 영어-중국어 전용 병렬말뭉치 분석도구를 개발해 병렬말뭉치의 최대 고민 중 하나인 원문과 번역문의 정렬 문제를 말끔히 해결했다. 릭순(Lixun 2001)은 자신이 개발한 병렬말뭉치 분석도구를 한국어와 일본어에도 확대 적용할 수 있을 것이라는 가능성을 언급하였는데, 그의 주장대로 한글-외국어 전용 검색, 분석 도구를 개발한다면 병렬말뭉치 연구 활성화에도 도움이 될 것이다.

4.3 병렬 검색(Parallel Search) 기능

그림 6 Parallel Search 검색 KWIC 결과 화면

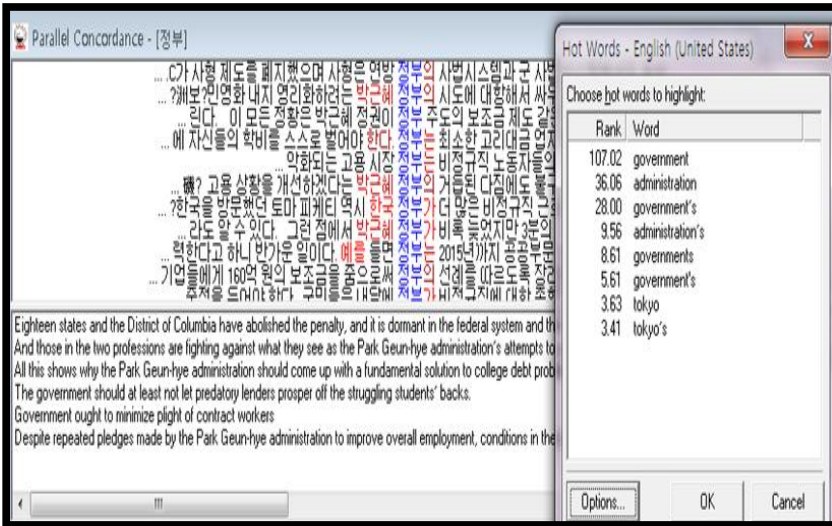


단일 말뭉치 분석도구와 달리 두 가지 이상의 언어를 함께 검색할 수 있는 ‘병렬 검색(parallel search)’ 기능은 병렬말뭉치 도구에서만 볼 수 있다. 위 그림 6은 이 기능을 이용해 영어의 ‘government’와 한글의 ‘정부’를 검색한 결과를 보여준다. Keyword in Context(KWIC) 형식으로 결과를 보여주고 있어서 학습자들이 영어와 한글의 검색어가 실제 문장에서 어떻게 사용되고 있는지를 검색어 앞, 뒤 환경을 보면서 파악할 수 있다는 장점이 있다.

4.4 Hot Words 기능

ParaConc의 독특한 기능 중 하나가 입력한 검색어의 가장 적절한 번역어를 말뭉치에서 찾아서 통계적으로 가장 가능성 높은 등가어(equivalent)를 찾아주는 기능이다. 이를 ‘Hot words’라고 부르는데, 아래 그림 7은 이 기능을 이용해 한글의 ‘정부’에 대응될 만한 영어의 등가어를 설정한 범위 내에서 8개 후보를 보여주고 있다.

그림 7 Hot Words 분석 결과 화면



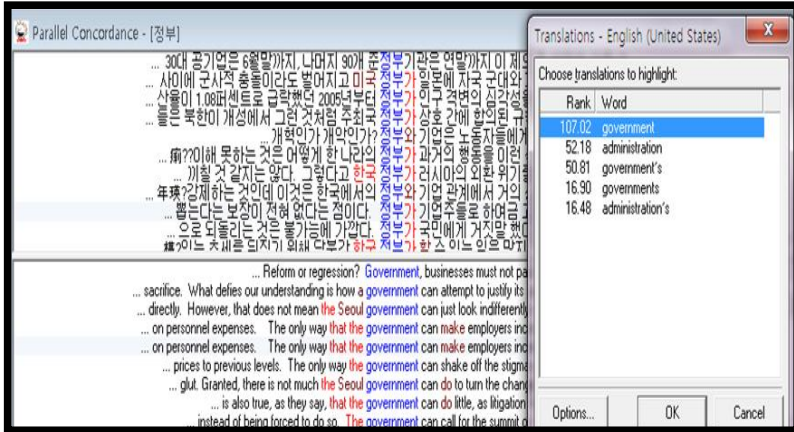
‘Hot words’를 통해 추출되는 번역 대응어 후보들은 통계적으로 출현빈도가

예상되는 빈도수 보다 높게 나타나는 경우에 해당한다. 이러한 ‘Hot words’ 기능을 이용하면, ‘정부’라는 단어가 영어로 ‘government’ ‘administration’ ‘tokyo’ 등으로 번역이 가능하다는 것을 알 수 있다. 물론 컴퓨터에서 통계적으로 처리한 결과 값이고 이는 어디까지나 가능성 높은 등가어를 보여주는 차원으로 이해하고 사용해야 한다. 여기서 마지막의 ‘tokyo’의 경우는 정부가 들어 간 문장에서 ‘정부’라는 단어를 영어로는 한 나라의 수도 명칭으로 해당 국가의 행정부를 대신하는 경우에 해당하는 경우로 설명될 수 있다. ParaConc의 이러한 독특한 ‘Hot words’ 기능을 효과적으로 이용하면 원문의 단어에 상응하는 번역 대응어로 구성된 어휘목록을 작성할 수 있다. 예를 들어 특정분야에서 사용되는 전문용어 목록 작성을 할 때, 병렬말뭉치와 병렬말뭉치 전용 분석도구를 사용하면 기존의 사전 등에서는 찾아 볼 수 없는 어휘목록 구성이 가능하다. 이런 점을 이용해 특정 분야의 어휘목록을 ‘원어-번역어’ 형식으로 추출해 정리할 수 있는데 예를 들어 의학 문서를 말뭉치로 구축해서 의학 분야 용어 추출을 위해 ParaConc의 Hot words 기능을 사용한 경우도 있다(Ndhlovu 2016). 이러한 노력은 용어목록 작업을 할 때 아주 유용하며, 이전에는 번역사들이 수작업으로 정리하면서 많은 시간을 들여야 했던 것과 달리 손쉽게 번역 전 용어정리와 같은 작업을 할 수 있다.

4.5 Translations 기능

Hot Words 기능과 유사하게 ‘Translations’ 기능을 사용해도 검색어와 가장 적절한 번역 등가어를 찾아준다. 아래 그림 8은 한글의 ‘정부’ 대응하는 영어의 등가어로 후보 5개를 지정한 결과로 ParaConc에서 분석한 후보를 보여주고 있다.

그림 8 Translations 분석결과 KWIC 화면



ParaConc에 추출된 검색어의 대응 번역어 후보 중 하나를 하단 그림 8의 화면 하단에서 클릭하면 KWIC에서 보다 상세한 문장을 볼 수 있는데, 이를 상단의 원문에서도 같은 문장을 추출해서 아래와 같이 정리하면, 원문과 번역문에서 특정 어휘 또는 어구가 어떤 식으로 번역이 가능한 지를 파악하는데 도움이 된다.

원문 1: **정부는** 선수 지원을 강화하고 동계 종목에 국민 관심을 유도하는 방안 등 구체적인 계획을 마련해야 한다.

번역문 1: **The government** should come up with specific plans, including supporting athletes and promoting ways to arouse public interest in winter sports.

원문 2: 또한 정부는 이런 결과를 그대로 받아들인 상태다.

번역문 2: It seems that **the Korean government** accepted IPK's report with few questions.

원문 3: **정부는** 이제부터 성주 주민들의 불만사항에 제대로 대응하기 위해 충력을 다해야 한다.

번역문 3: The Park Geun-hye administration should make every effort to address the complaints of Seongju from here onwards.

원문 4: 박근혜 정부는 소를 잃은 뒤에도 외양간을 고치지 못했다.

번역문 4: The Park administration has failed to mend the barn even after the horses had bolted.

원문 5: 미국 정부가 재부상하는 중국을 견제하고 자국의 군산 복합체의 필요를 만족시키기 위해 한반도에서 일정한 긴장을 유지하는 쪽을 선호하는 것이 아닌가 하는 의혹을 확산시킨다.

번역문 5: It will further spread suspicion that the U.S. administration prefers some tension being maintained on the Korean Peninsula both to keep a resurgent China in check and satisfy the needs of the US military-industry complex.

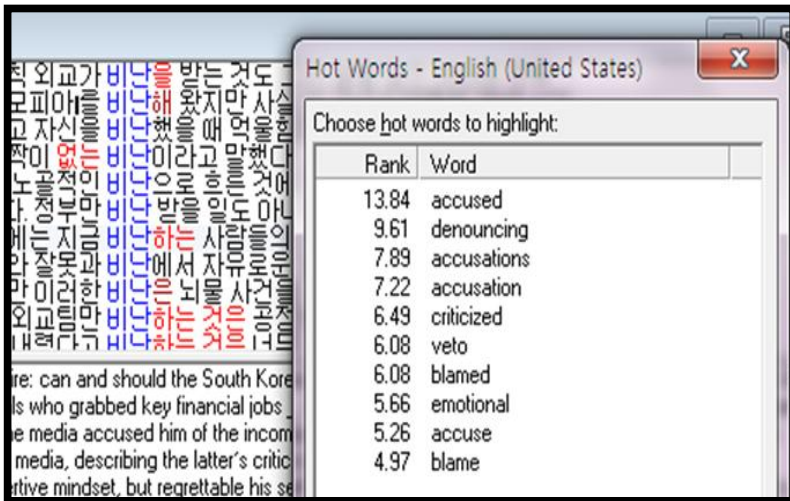
바로 위에 제시한 이러한 원문과 번역문 비교, 분석활동은 번역을 배우는 학생들에게 아주 유용한 수업자료로 활용이 가능하다. 소위 'data-driven learning (DDL) (Johns 1991)을 통해 학습자들이 병렬말뭉치를 상호 비교, 분석하면서 영어와 한국어 어휘, 문법 등의 유형을 익히고 이를 통해 모국어 및 외국어에 대한 이해도를 제고할 수 있다. 이러한 귀납적 학습활동 기반의 병렬말뭉치를 활용한 번역교육의 가능성은 앞으로도 많은 연구와 노력이 필요한 영역이라 생각된다.

4.6 Concordancing을 이용한 등가어 찾기 수업활동

위에서 설명한 ParaConc의 기능을 이용해서 특정 표현의 번역 후보들을 학습자들이 직접 찾아보는 활동을 정리해본다. 말뭉치 언어학은 컴퓨터를 활용한 데이터 정리 및 분석을 하기 때문에 말뭉치 분석도구의 활용이 무엇보다 절대적이다. 그래야 방대한 양의 데이터를 빠르고 정확하게 분석할 수 있기 때문이다. 이러한 분석도구를 concordancer 또는 concordancing tool이라고 부른다. 혼

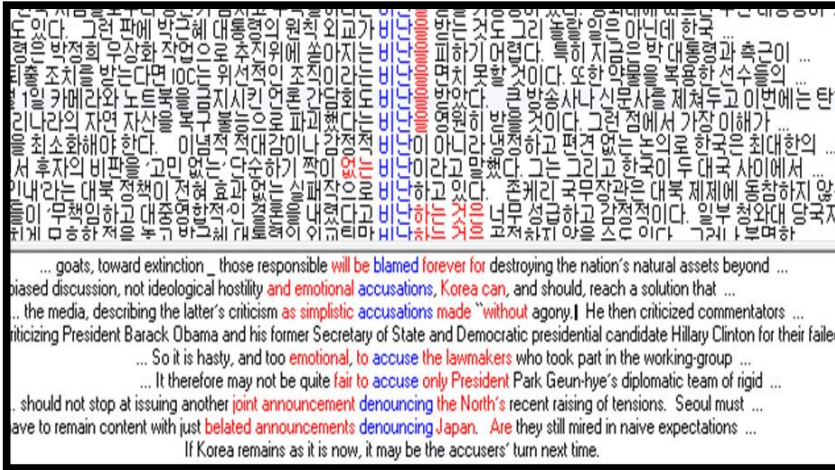
히 수집한 텍스트에서 어휘나 어구의 빈도수를 기준으로 추출해 컴퓨터 화면에 보여주는 그림 5, 6, 7 그리고 8의 한줄 한 줄을 concordancing line이라고 한다. 번역사 또는 학습자가 컴퓨터를 통해 분석한 결과를 보다 심도 있게 분석해서 정리하는 활동은 수동적인 교실활동에서 벗어나 ‘language detective(Johns 1997)’ 즉 말뭉치를 자세히 관찰하고 분석하는 Sherlock Homes와 같은 탐정이 되어 적극적으로 학습활동 참여를 가능하게 한다. 아래 그림 9는 ‘비난’을 ‘정규식 표현(regular expressions)’ 검색을 통해 ‘비난’이 들어가는 원문 한글 텍스트의 표현을 추출한 후 ‘Hot words’ 기능을 사용해 번역 후보 목록을 추출한 결과를 보여주고 있다.

그림 9 검색어 ‘Hot words’를 이용해 ‘비난’의 번역 후보군 추출 화면



이렇게 추출된 후보 목록을 기준으로 아래 그림 10과 같이 KWIC 검색결과를 만들어서 원문과 번역문의 비교를 간단하게 만들 수 있다.

그림 10 '비난'의 번역 후보군 KWIC 검색 결과



위 그림 10에 추출된 한글과 영어의 '비난'과 관련된 어구는 총 40개인데, 이 중에서 한글과 영어의 비난 관련 단어 및 표현을 <표 4>와 같이 정리할 수 있다.

<표 4> '비난'의 한글과 영어의 표현 정리

한글	영어
~을 비난하다	accuse ~; blame ~; condemn~; criticize ~; denounce; lash out at~; rebuke~;
~으로 인한 비난에서 자유로울 수 없다	be not free from blame for ~
~으로 인해 비난을 받다	be blamed be under fire come under fire get the blame for ~
~에 대한 비난을 피하다	avoid blame for~
~에 대한 비난	accusation of ~
비난 받아 마땅하다	deserve criticism

한글 1) 김 특사는 북한이 ‘전제조건 없는 탐색적 대화’에 응하지 않는다고 **비난했지만** 오래 지연된 과정에 또 하나의 준비 단계를 추가하는 것이 좋은 생각인지는 의심의 여지가 있다.

영어 1) Kim **accused** the North **of** not responding to “exploratory talks without preconditions,” but it is open to question whether adding one more preparatory stage to the long-delayed process is a good idea.

한글 2) 정부만 **비난 받을** 일도 아니다.

영어 2) The government is not alone in **deserving blame**.

한글 3) 한 주류 의원이 그를 ‘공갈꾼’으로 **비난하자** 전자는 개인적 치욕에 분노하면서 문을 박차고 나가버렸다

영어 3) A mainstream lawmaker **criticized him of** being a “blackmailer,” forcing the former to stomp out of the room, enraged at the personal humiliation.

한글 4) 그런 판에 박근혜 대통령의 원칙 외교가 **비난을 받는 것도** 그리 놀랄 일은 아닌데 한국 지도자가 북한은 비핵화 문제에서 그리고 일본은 “위안부” 문제에서 더 “진정성”을 보이라고 계속 요구할 수가 있으며 또 그래야 할까 하는 점이다.

영어 4) It is small surprise then Park’s principle-first diplomacy has **come under fire**: can and should the South Korean leader continue to demand that North Korea and Japan show more “sincerity,” concerning the issues of denuclearization for Pyongyang, and “the comfort women” for Tokyo?

한글 5) 그러나 그렇다고 해서 한국 외교관들이 뒤늦게 **일본을 비난하는** 성명을 내는 정도로 만족할 필요는 없다.

영어 5) Of course Seoul lags far behind Tokyo in economic and other areas, but that should not necessarily mean Korean diplomats have to remain content with just belated announcements **denouncing Japan**.

한글 5) 박대통령은 핵심 금융 직을 꿰어 찬 전직 재무부 관료들인 “모피아”를 비난해 왔지만 사실상 그녀가 임명한 것이나 다름없는 사람들 중 일부는 금융계에 경험조차 없다.

영어 5) Park lashed out at the “Mofia” _ former Ministry of Finance (MoF) officials who grabbed key financial jobs _ but some of her (virtual) appointees do not even have experience in the financial world.

한글 6) 11월 4일 발표된 대국민 사과와 그 후 정세균 국회의장 면담이 역풍을 맞은 원인은 권력 이양에 대한 명확한 메시지가 결여됐기 때문이다.

영어 6) Her second apology delivered Nov. 4 and a subsequent meeting with National Assembly Speaker Chung Sye-kyun were heavily criticized because she was vague about delegating her authority to a new prime minister.

한글과 영어 말뭉치에서 해당하는 표현을 찾아 문장 단위로 정리를 해서 비교하면, 실제 원문이 어떻게 번역이 이뤄졌는지를 알 수 있고, 이를 통해 동일한 단어, 어구 등이 실제 번역에서 다양하게 번역될 수 있다는 점이 부각될 수 있고 동사와 연결되는 전치사 또는 명사와 어울리는 동사 표현 등의 연어 (collocation)에 대한 인식을 제고하는데도 의미가 있다. 이러한 concordancing 활동은 단순히 사전을 검색해서 해당 어휘나 표현을 찾아 적는 것보다 훨씬 학습효과가 뛰어나다고 할 수 있다.

5. 결론

5.1 결론

많은 학자들이 병렬말뭉치의 번역 교육이나 연구 및 대조언어 분석 연구에 활용가치가 높다고 주장해왔다(김성식 2012; 박명수 2017; 정호정 2003; Aijmer & Altenberg 1996; Bernardini 1997; Bernardini & Zanettin 2004; Laviosa

1998; McEnergy & Xiao 2002; Xiao & McEnergy, 2002; Zanettin 1998). 이러한 주장이 수없이 되풀이 되고 있지만, 실제 번역교육, 외국어교육, 번역연구, 외국어교육 연구 등에 병렬말뭉치가 활용되는 사례는 다른 분야 연구보다 상대적으로 부족한 실정이다. 양적인 측면에서 연구가 부족하다는 것 이외에, 우리나라에서 발표된 연구를 토대로 병렬말뭉치에 대한 연구를 분석해본 결과(표 2 참조), 두 가지 문제점이 발견되었다. 첫째, 말뭉치 정의의 재정립이 필요하다. 말뭉치란 컴퓨터로 읽어낼 수 있는 방대한 양의 텍스트를 일정한 목적과 원칙에 입각해 수집한 것을 의미한다. ‘방대한 양’이라는 정의는 연구 목적에 따라 말뭉치 규모가 달라질 수 있고, 최근 들어 특수목적 영어(English for Specific Purposes)와 의학, 과학, 기술, 법률 등과 전문영역 번역과 관련해서 특수목적 코퍼스(specialized corpus)가 사용되고 있기 때문에 규모는 크게 문제되지 않을 수 있다. 적절한 규모의 말뭉치에 대한 질문의 답은 목적이 무엇인가에 따라 달라진다(박명수 2017). 선행연구에서 말뭉치에 대한 정의와 관련해 재정립이 필요한 부분은 ‘컴퓨터가 인식할 수 있는(machine-readable)’이라는 부분이다. 저작권 문제 등으로 인해 말뭉치 자료 수집의 한계가 있으나 이것이 컴퓨터 기반이 아닌 종이 자료를 모아 놓아도 말뭉치라 할 수 있다라는 정당성을 주는 것이 아니다. 번역학에서는 말뭉치의 정의를 특정 기준에 따라 수작업으로 수집해서 분석할 수 있는 비교적 작은 규모의 텍스트 모음이다(Laviosa 2002)라는 주장을 선행연구에서 주로 인용했다. 그러나 이러한 정의는 특정 작가, 특정 번역사, 특정 시기, 특정 장르에 국한된 연구로 제한하거나 컴퓨터 기반의 말뭉치 구축이 불가능한 경우로 국한되어야 한다. 라비오사(Laviosa)가 정의를 내린 시점이 인터넷과 컴퓨터 기반 언어 데이터 분석 기술의 초창기였다는 점을 감안한다면 이제 번역학에서도 말뭉치에 대한 정의를 재정립해보는 것도 필요하다.

두 번째 문제는, 말뭉치를 분석할 수 있는 도구의 부재이다. 대부분의 선행연구는 이러한 이유로 인해 말뭉치를 수집한 후 분석도구를 사용하지 않고 수작업으로 분석을 진행했다. 유현경·안예리·양수향(2007)은 한영 병렬말뭉치를 분석할 마땅한 검색도구가 개발되어 있지 않은 실정에 대해 언급한 바 있다. 국립국어원에서 제작한 한영 병렬말뭉치 분석도구가 있기는 하나 이미 구축되어 있는 말뭉치만 분석할 수 있고, 연구자가 직접 구축한 외부 말뭉치를 분석하는 것은 불가능하기 때문에 현재로서는 범용적으로 활용할 수 있는 병렬말뭉치

분석도구는 찾아볼 수 없다. 본 연구에서 소개한 ‘ParaConc’ 이외에 비교적 최근에 개발된 ‘AntPConc⁴⁾’ 등이 공개되어 있으나, 아직까지 우리 한글과 병렬로 구축된 외국어를 이러한 도구로 분석을 시도한 예는 찾아보기 어렵다. 번역학 연구 분야가 진일보하려면 컴퓨터 기술 분야와 협력이 절실한 대목이다. 연구자 개인이 이러한 분석도구를 개발하려면 비용과 시간이 만만치 않기에 국가적 차원의 지원과 노력이 요구된다. 한글 전용 병렬말뭉치가 개발된다면 말뭉치의 규모도 크게 확장할 수 있고 폭넓고 심도있는 연구도 활성화 될 것이다.

5.2 교육적 함의

본 연구는 영·한 병렬말뭉치 분석을 위해 ‘ParaConc’의 활용 가능성을 분석했다. 앞서 언급한대로 한글 전용 병렬말뭉치 분석도구의 부재로 인해 말뭉치 기반 번역 연구 및 교육에 제대로 활용된 사례가 부족하다. 과거에는 원문과 비교해 번역문이 질적으로 수준이 떨어진다는 이유로 말뭉치를 기반으로 한 번역 연구를 금지해야 한다는 주장도 있었다(Hansen-Schirra & Teich 2009). 그러나 번역학이 독립된 학문으로 자리를 잡고 다양한 연구와 교육이 이뤄지는 현실에서 볼 때 이러한 관점은 더 이상 유효하지 않다. 무엇보다 말뭉치를 활용해 번역 교육의 가능성은 무궁무진한데 실용적인 번역 연구 및 교육 관점에서 말뭉치는 언어 및 문화적 인식을 제고시킬 수 있다(박명수 2014; Bernardini 1997; Hunston 2002). 본 연구에서 소개한 ‘ParaConc’ 기능 중에서 ‘Hot words’를 사용해 추출된 콘코던스 라인(concordance line)을 분석하는 활동은 번역 교육에서 시도해볼 수 있는 DDL 기반 수업활동이 될 수 있다. 이미 DDL 기반 수업활동이 외국어 교육 분야에서 다양하게 활용되고 있지만, 번역 교육 분야에서는 콘코던스 라인을 이용하거나 KWIC 검색 결과를 활용한 번역 교육 사례는 상대적으로 부족하다. 병렬말뭉치의 콘코던스 라인을 이용하면 (그림 8과 10 참고) 학습자들이 영어 또는 한글의 특정 어휘나 표현이 번역문에서 어떤 식으로 번역이 되었는지를 직접 찾아볼 수 있기 때문에 용어, 표현 등을 단순히 사

4) AntPConc는 일본의 와세다 대학의 Lawrence Anthony 교수가 제작한 병렬말뭉치 분석도구로, <http://www.laurenceanthony.net/software/antpconc/>에서 무료로 다운로드해서 컴퓨터에 설치해 사용할 수 있다.

전에서 찾는 활동보다 더 적극적인 학습참여를 유도할 수 있고, 학습자 스스로가 관찰과 분석을 통해 필요한 어휘, 표현, 용어 등을 찾아보는 발견학습(discovery learning)도 가능하다. 이러한 병렬말뭉치 활용 학습 활동을 하면 보다 더 정교하게 번역을 하는데 도움이 될 수 있으므로 병렬말뭉치는 교육적 가치가 높다(Williams 1996).

본 연구는 병렬말뭉치 분석도구인 'ParaConc'를 이용해 영-한 병렬말뭉치를 분석해 번역 연구 및 교육의 활용 가능성을 찾아보고자 했다. 이러한 콘코댄서(concordancer)를 사용해 연구자들이 수작업으로 말뭉치를 분석하지 않고 보다 수월하게 많은 양의 데이터를 분석하는데 미력하나마 도움이 되기를 바란다. 마지막으로 'ParaConc'가 한글 원문이나 번역문을 완벽하게 분석하는데 제한적이기 때문에 향후 한글 전용 병렬말뭉치 분석도구가 개발되어 번역 연구 및 교육이 한 단계 성장하는 계기가 마련되기를 희망한다.

참고문헌

- 김성식 (2012) 「ParaConc를 활용한 번역연구」, 『글로벌 영어교육학회 2012 국제학술대회 발표논문집』 전주교육대학교 2012년 11월 24일.
- 김한샘 (2009). 「한국어 관용구의 영어 번역 양사 -한영병렬 말뭉치를 기반으로-」, 『언어와 문화』 5(3): 241-259.
- 김홍규, 정태구, 김정숙 (2000) 「한-영 병렬 코퍼스의 설계·구축 및 응용 방안 연구」, 『한국어학』 11: 23-71.
- 민경모 (2010) 「병렬말뭉치의 개념 및 구조에 관한 몇 문제」, 『언어사실과 관점』 25: 41-70.
- 박명수 (2017) 『코퍼스 언어학의 이해와 실제』 서울: 한빛문화.
- 박명수 (2014) 「번역교육을 위한 코퍼스기반 용어추출 방법」, 『번역학연구』 15(1): 111-134.
- 박정열, 차정원 (2008) 「한영 병렬 코퍼스 구축을 위한 하이브리드 기반 문장 자동 정렬 방법」, 『말소리』 68: 95-114.
- 서상규 (2009) 「국어 특수 자료 구축의 성과와 전망」, 『새국어생활』 19-1 (봄),

- 국립국어원.
- 서정목 (2010) 「한·영 병렬 코퍼스의 설계·구축 및 응용 방안 연구」, 『한국어학』 11: 23-71.
- 송상현 (2010) 「한국어 과거시제 표지의 병렬 말뭉치 기반 분석」, 『언어정보』 20: 75-104.
- 신자영 (2010) 「병렬 코퍼스 및 학습자 코퍼스를 이용한 중간언어 연구법」, 『언어사실과 관점』 25: 71-87.
- 신자영, 원미진 (2011) 「한국어교육을 위한 병렬말뭉치의 대조 주석 모형 개발 방안」, 『언어와 정보 사회』 14: 97-124.
- 조의연 (2009) 「병렬 말뭉치에 기반한 번역학 연구」, 『번역학연구』 10(2): 207-246.
- 조준형 (2012) 「병렬코퍼스에서 맥락 탐색의 의미와 한계」, 『번역학연구』 13(5): 223-246.
- 주향아, 조미희 (2010). 「신문기사 한영 병렬 코퍼스를 통해 본 번역 단위 -시장, 기업을 중심으로-」, 『언어사실과 관점』 25: 161-188.
- 정호정 (2003) 「코퍼스 중심의 번역학 연구」, 『번역학연구』 4(2): 89-115.
- 최승권, 김영길 (2010) 「번역 말뭉치로부터 추출한 어휘 번역패턴의 의미 분류와 자동번역시스템에의 활용」, 『번역학연구』 11(3): 277-301.
- 최정아 (2003) 「병렬 말뭉치를 통한 한국어-영어의 번역 단어수 연구」, 『번역학연구』 4(2): 89-115.
- Aijmer, Karin, Altenberg, Bengt., and Johansson, Mats (1996) *Languages in contrast: Papers from a Symposium on text-based cross-linguistic studies*, Lund: Lund University Press.
- Anthony, Lawrence. (2014) AntPConc (Version 1.1.0) [Computer Software]. Tokyo, Japan: Waseda University.
- Barlow, Michael (2001) *ParaConc* [Computer software]. Houston, TX: Athelstan.
- Bernardini, Sylvia (1997) A 'trainee' translator's perspective on corpora. Paper presented at Corpus use and learning to translate held at Bertinoro Nov. 1997. Retrieved November, 2016 from <http://www.sslmit.unibo.it/introduz.htm>.
- Bernardini, Sylvia and Zanettin, Federico (2004) 'When is a universal not a

- universal? Some limits of current corpus-based methodologies for the investigation of translation universals', in Anna Mauranen, & Pekka Kujamäki (eds.) *Translation universals: Do they exist?* Amsterdam: John Benjamins, 51-62.
- Hansen-Schirra, Silvia and Teich, Elke (2009) 'Corpora in human translation', in Anke Lüdeling, Merja Kytö, and Tony McEnery (eds) *Corpus Linguistics: An International Handbook*. Berlin and New York: de Gruyter, 1159-1175.
- Johns, Tim (1991) 'From printout to handout: Grammar and vocabulary teaching in the context of data driven learning', *English Language Research Journal* 4: 27-45.
- (1997) 'Contexts: The background, development and trialling of a concordance-based CALL program', In Anne Wichmann, Steven Fligelstone, Tony McEnery, and Gerry Knowles (eds) *Teaching and language corpora*, Longman, 100-115.
- Lixun, Wang (2001) 'Exploring parallel concordancing in English and Chinese', *Language Learning & Technology*, 5(3): 174-184.
- Laviosa, Sara (1998) 'The corpus-based approach: a new paradigm in translation studies', *Meta* 43(4): 474-479.
- (2002) *Corpus-based translation studies: Theories, findings, applications*. Amsterdam and New York: Rodopi.
- McEnery, Anthony and Xiao, Zhonghua (2002) 'Domains, text types, aspect marking and English-Chinese translation', *Journal of Languages in Contrast* 2(2): 211-31.
- (2004) 'Character encoding in corpus construction', in Martin Wynne (ed) *Guide to good practice*. Oxford: Oxford University Press, 47-58.
- (2005) 'Parallel and comparable corpora: What are they up to?', in *Proceedings of Incorporating Corpora: Translation and the Linguist. Translating Europe. Multilingual Matters*, Clevedon, UK.
- Ndhlovu, Ketiwe (2016) 'Using ParaConc to extract bilingual terminology from

parallel corpora: A case of English and Ndebele', *Liberator* 37(2). a1278. <http://dx.doi.org/10.4102/lit.v37i2.1278>.

Williams, Ian. (1996) 'A translator's reference needs: dictionaries or parallel texts?', *Target* 8(2): 277-99.

Xiao, Zhonghua and McEnery, Anthony (2002) 'A corpus-based approach to tense and aspect in English-Chinese translation', Paper presented at International Symposium on Contrastive and Translation Studies between Chinese and English. Shanghai 2002.

Zanettin, Federico (1998) 'Bilingual comparable corpora and the training of translators'. *Meta* 43(4): 616-630.

[Abstract]

Investigation into English-Korean Parallel Corpus with ParaConc

Park, Myongsu
(Sangmyung University)

This study aimed at 1) analyzing the parallel corpus-based research published in major Korean journals related to translation, and 2) attempted to use ParaConc dedicated to analyzing parallel corpora to investigate into the English-Korean parallel corpus. It was found that the parallel corpus-based translation is still under-researched with only 23 studies published from 2000 to 2017. The lack of a concordancer for parallel corpora was cited as one of the main reasons of such fewer parallel corpus-based translation studies. As part of efforts of encouraging the use of a parallel corpus and a concordancer dedicated to parallel corpora, the present study built an English-Korean parallel corpus with editorials published in major vernacular English newspapers and then attempted to use ParaConc to see how the tool can be used for studying translation issues, educating English language learners and further contributing to the translation studies. The suggested methods can serve as more efficient and engaging ways of teaching and learning in English-Korean translation classrooms. As there is no concordancer dedicated to analyzing English-Korean or Korean-English parallel corpora, a nation-wide research support is indispensable to develop a concordancer dedicated to Korean-English/English-Korean corpora, which definitively contribute to the translation studies in the long run.

▶ Key Words: corpus, concordancer, parallel corpus, ParaConc

박명수

상명대학교 글로벌지역학부 영어권지역학 부교수

myongsu@smu.ac.kr

관심분야: 통·번역, 코퍼스언어학, 코퍼스기반 번역학, 영상번역

논문투고일: 2017년 10월 29일

심사완료일: 2017년 12월 12일

게재확정일: 2017년 12월 13일