

학술대회발표논문집

2012년 대한언어학회 가을 학술대회

- 주제: 언어학 연구의 실제적 활용 - 이론, 교육 및 표기법을 중심으로
- 장소: 광주 전남대학교 진리관(경영대학과 인문대 1호관 사이)
- 일시: 2012. 10. 20(토) 09:00 ~ 18:00
- 등록비: 1만원(점심 및 프로시딩스 포함)
- 주최: 대한언어학회
- 주관: 전남대학교 영어교육과
- 후원: 한국연구재단, 전남대학교

- * 이 발표논문집은 2012년도 정부재원(교육과학기술부)으로 한국연구재단의 지원을 받아 전자발간 되었음
- * This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government.

과학저널에 나타난 어휘다발의 사용 연구

이성민
(부경대학교)

1. 서론

언어학자들은 지난 한 세기동안 언어의 패턴에 관한 연구에 관심을 가져왔다. 언어(collocation)란 용어가 Palmer(1933)에 의해 연속되는 두 단어나 그 이상의 단어(a succession of two or more words)로 정의되었고, Firth(1951)는 단어가 다른 단어와 습관적으로 나타나는 현상을 설명하고자 하였다. 하지만 20세기 후반기에 코퍼스언어학(corpus linguistics)의 성장으로 통계적으로 높은 빈도수를 가지며 자연스런 문맥 속에 나타나는 언어의 확장된 형태에 관한 연구가 본격화 되었다. Hakuta(1974)의 선 조합패턴(pre-fabricated patterns), Decarrico(1992)의 어휘구조(lexical phrases), Scott(1999)의 말무리(clusters), Biber *et al.*(1999)의 어휘다발(lexical bundles), Wray(2002)의 정형화된 연속체(formulaic sequences), O' keefe *et al.*(2007)의 말뭉치(chunks), Granger & Meunier(2008)의 어구 연쇄(Phraseology)등 학자마다 다양한 용어를 사용하여 왔다.

어휘다발의 연구방향은 크게 두 가지로 나누어 볼 수 있다. 첫째, 원어민 코퍼스를 대상으로 한 연구이다. 대표적 연구로서 Biber *et al.*(1999)에서는 대화(conversation)와 학술작문(academic prose)에서 발견된 어휘다발의 비교를 통해 사용역(register)에 따른 차이를 보여주었고, 후속연구인 Biber, Cornard & Cortes(2004), Biber & Barbieri (2007)에서 대학에서 사용되는 여러 영역의 문어(written)와 구어(spoken)를 대상으로 어휘다발의 특성을 관찰하였다. Cortes(2004)는 역사와 생물학분야의 논문과 대학생 보고서를 비교하여 학문분야별, 사용자 수준별 어휘다발이 차이가 있음을 밝혔고, Hyland(2008a)은 전자공학, 생물학, 경영학, 언어학분야의 출간된 논문과 석.박사 논문으로 구성된 코퍼스를 대상으로 각 장르별로 사용하는 선호하는 어휘다발이 다름을 보여주었다.

둘째, 학습자 코퍼스(learner corpus)을 대상으로 한 비교연구다. Granger(1998)은 3백만 단어 규모의 비원어민 작문으로 구성된 International Corpus of Learner English(ICLE)연구에서 비원어민의 어구 연쇄 사용능력(phraseological skill)이 부족하며, 원어민에 비해 적은 어휘다발을 사용한다고 하였고, De Cock(2004: 243)은 비원어민의 과도(overuse), 과소(underuse), 오류(misuse) 현상을 보고하였다. 영어가 모국어인 학자와 대학생, 중국어가 모국어인 비원어민 대학생의 학술작문을 비교한 연구에서 Chen & baker(2010)는 비원어민이 구어체사용, 문법오류, 일반화 경향 그리고 헤지(hedge)표현 사용부족을 보인다고 보고하였다.

국내 연구로는 어휘다발의 사용양상이 국가, 시대, 장르의 3가지 요소에 따라 어떻게 달

라지는지 연구한 박성우(2009)와 한국인 대학생 영어 학습자 코퍼스에서 나타난 정형화된 연결어구의 특성을 연구한 이은주(2009) 그리고 영어 구어 담화 코퍼스인 Michigan Corpus of Academic Spoken English(MICASE)를 활용하여 각 학문분야에 나타나는 정형어구를 추출하고, 구조적 그리고 기능적 범주에 따른 특징을 조사한 권예은·이은주(2010)의 연구가 있었다.

본 연구의 목적은 학술장르(academic genre)에서 비 원어민 학습자로서 한국인이 사용한 어휘다발을 원어민과 비교 분석함으로써 그 언어적 차이와 특성을 살펴보고 나아가 발생 원인을 파악하는데 있다. 이를 위해 생명과학 관련 논문을 선정하여 연구를 진행하였으며 그 이유는 다음과 같다. 첫째, 어휘다발의 사용은 학문영역별로 사용빈도(frequency), 구조(structure), 의미(functions)에서 중대한 차이를 보인다(Cortes 2004; Biber 2006; Hyland 2008a). 따라서 여러 분야를 합치게 되면 원어민과 비원어민의 특성뿐만 아니라 장르의 영향을 받지 않을 수 없기 때문에 한 가지 분야로 한정하였다. 둘째, 생물학코퍼스를 대상으로 한 선행연구(Cortes 2004; Hyland 2008a; Danica 2011)와 비교 연구가 가능하다. 셋째, 생명과학 분야는 구체적인 실험을 중요시하며 공식화된 논의방식을 사용하기 때문에 실험연구를 중심으로 하는 전형적인 과학보고서의 특성을 공유하고 있으므로, 본 연구에서 추출한 어휘다발은 자연과학(hard science)분야 논문작성의 교수·학습 자료로 활용이 가능하다. 넷째, 특정 학문영역에서 관습적으로 사용되는 어휘다발의 사용능력은 저자의 사용능력을 평가하는 근거가 될 수 있다(Cortes 2004: 398). 따라서 한국과학자 어휘다발의 특성을 파악하면 해당 분야에서 한국인의 학술작문이 얼마나 영어다운지 확인할 수 있게 되는 것이다.

본 연구의 구성은 다음과 같다. 2장에서는 핵심개념과 분석 대상 및 방법을 3장에서는 어휘다발 분석결과 어휘다발의 구조별로 나누어 분석하고, 4장 결론에서는 연구결과를 요약한다.

2. 분석 대상 및 방법

2.1. 어휘다발

어휘다발(lexical bundle)이라는 용어는 Longman Grammar of Spoken and Written English(LGSWE)에서 처음 사용되었는데, Biber *et al.*(1999: 990)은 자연스런 문맥에서 함께 나타나는 연속된 단어형태(“sequences of word forms that commonly go together in natural discourse”)라고 정의하였다. 어휘다발은 언어(‘collocation’)의 확장된 형태로 볼 수 있으며 반드시 관용어처럼 쓰이거나 구조적으로 완전하지 않지만 담화를 구성하는 특정한 의미를 가진 단어조합으로 볼 수 있다. 구체적으로는 연속된 3단어 이상에서 사용되는 빈도가 백만 단어 당 10회 이상 그리고 5개 이상의 서로 다른 텍스트에서 출현하는 경우를 어휘다발로 보며, in terms of, as a result of, on the other hand, in the context of를 예로 들 수 있다.

2.2. 사용 코퍼스

한국인과학자코퍼스는 포항공대 생물학연구정보센터¹⁾ 홈페이지에서 외국 생명과학관련 주요학술지에 투고한 한국과학자들의 영어 논문을 대상으로 하였다. 수집방법 및 코퍼스작성 절차는 다음과 같다. 2010년 1월부터 2012년 1월까지 웹페이지에 게시된 논문가운데 50편을 선정하였다. 그런 다음 PDF나 웹페이지 형식으로 제공되는 원 논문내용을 각각 텍스트로 저장하였다. 동시에 엑셀파일에 저자이름, 논문이 게재된 학술지명, 웹페이지 게재 일자를 기록한 메타 데이터를 작성하였다. 이 코퍼스는 모두 50개 텍스트로 구성되었으며, 누적 단어수는 약 270,969개, 평균 단어수는 4,098개, 명칭은 Korean Life Science(KLS)코퍼스로 정하였다.

참조할 원어민과학자코퍼스는 총 1천 4백만 단어로 구성된 공개된 ANC(American National Corpus)의 하위 부류인 Biomed 코퍼스²⁾를 대상으로 하였다. Biomed 코퍼스는 미국인저자가 작성한 논문만으로 구성되어 있으며 누적 단어수는 3,480,722개, 평균 단어수는 3,839개 이다.

KLS코퍼스와 Biomed코퍼스가 비교 가능한 이유는 두 코퍼스가 모두 생물학과 그 응용분야인 의학논문을 모은 코퍼스로서 생명과학논문으로 분류 될 수 있기 때문이다. 실제로 KLS코퍼스 경우 저자의 50%이상이 의사였으며, 나머지 대학교수 및 전문연구원이었다. 마찬가지로 Biomed코퍼스 또한 생물학 저널인 BMC Biology와 의학저널인 BMC Medicine을 대상으로 2000 ~ 2003년 4년간 수집한 미국인저자가 작성한 논문으로 구성된 코퍼스이기 때문에 이 두 코퍼스의 비교를 통하여 정확한 언어 사용의 양상을 파악할 수 있으며 어휘다발의 사용과 관련된 통계적인 자료를 제공하여 줌으로써 비원어민 학습자로서 한국인과 원어민과학자가 보여주는 언어 사용의 패턴을 확인 할 수 있다.

2.3 분석도구 및 방법

본 연구에서 사용한 코퍼스 분석용 프로그램은 현재 전 세계적으로 가장 널리 쓰이고 있는 WordSmith Tools 5.0이다. 프로그램을 활용한 분석과정은 다음과 같다. 첫째, WordSmith Tools의 wordlist 기능을 활용하여 각 코퍼스에 대한 index³⁾자료를 산출하고, 둘째, 만들어진 index자료를 기반으로 하여, 상단compute메뉴의 cluster기능을 활용하여 3~4단어로 이루어진 어휘다발을 산출한다. 3~4단어로 이루어진 어휘다발만을 연구대상으로 선정한 이유는 이들이 가장 사용빈도가 높아 언어사용양상을 잘 관찰할 수 있으며(Biber et al., 1999: 994), 대다수의 연구가 3~4단어 어휘다발을 연구대상으로 하기 때문에 관련 선행연구와 비교가 가능하기 때문이다.

1) 현재 국가지정 생물학연구정보센터(Biological Research information Center)로 운영되고 있으며, 웹사이트 주소는 <http://bric.postech.ac.kr/>와 같다.

2) Biomed 자료는 2000~2003년 동안 Biomedcentral.com에서 수집하였는데, 이 홈페이지는 생물학과 의학전 분야에 걸친 논문을 무료로 제공하고 있다.

3) Index는 텍스트에 있는 모든 개별 단어의 위치를 파악하여 이를 바탕으로 개별 문장의 길이나 텍스트의 길이 또는 cluster기능을 사용할 수 있게 해 주는 기능이다(Scott, M., 2012).

2.4. 어휘다발 선별 및 삭제기준

각 코퍼스 어휘다발의 출현 빈도를 기준으로 삼았다. 어휘다발의 빈도수 산출범위는 특별히 정해진 바가 없으나 참고가 되는 연구를 기준으로 보수적(conservative) 내지는 상대적으로 높은(relatively high) 정도로 표현할 수 있다. 예컨대, Biber & Barbieri(2007: 267)처럼 만약 100만 단어당 40회 이상 출현한 4단어를 어휘다발로 인정하다고 하면 20회나 10회로 정한 다른 연구들에 비해서 더 보수적(conservative)하다고 말할 수 있다. 본 연구에서는 이상과 같은 연구결과를 참고하여 100만 단어당 30회 출현하는 연속된 3~4단어를 어휘다발로 설정하였다. 따라서 한국인과학자코퍼스의 크기가 27만 단어이므로 산출빈도는 8회가 되었고, 원어민과학자코퍼스는 340만 단어이므로 90회로 설정하였다.

출현빈도와 관련하여 어휘다발을 추출하기 위한 또 하나의 기준은 몇 개의 각기 다른 텍스트에서 나타나는 것을 대상으로 하는가이다. 이렇게 추가적인 제한을 하는 이유는 각기 다른 화자(speakers)나 저자(writers)에게서 공통적으로 사용되는 경우를 어휘다발로 보기 때문이며 또한 어휘다발의 과대한 증가(inflation)을 피하기 위함이다(Biber & Barbieri, 2007). Biber *et al.*(2004)은 개별 화자나 저자만의 특유한(idiosyncratic) 어휘사용⁴⁾의 경우를 제외하기 위하여 최소 5개 각각 다른 텍스트에서 사용된 경우만을 어휘다발 인정하였다. 이상의 연구방식을 고려하여, 본 연구에서도 최소한 5회 이상의 서로 다른 텍스트에서 나타난 경우를 어휘다발로 설정하였다.

마지막으로 고려해야 할 중요한 선별요인은 삭제기준이다. 실제로 컴퓨터 프로그램을 이용하여 기계적으로 3~4단어 어휘다발을 산출하게 되면, 높은 빈도를 기록하는 어휘다발 중에는 숫자를 포함한 데이터와 함께, 고유명사와 자주 결합하는 문맥의존표현(context-dependent expressions)이 리스트에 함께 나타난다(Chen & Baker 2010: 33). 따라서 본 연구에서는 생물학관련 전문용어나 특정 영역과 관련된 어휘다발을 수작업으로 일일이 제거하였다. 또한 관사(a, an the)로 끝나는 어휘다발 가운데 이미 다른 어휘다발의 부분임이 확인되었을 경우 생략하였는데, 예를 들면 in terms of(5), in terms of the(5)가 발견된 경우 불필요한 중복을 막기 위하여 in terms of만을 채택하였다.

3. 어휘다발 분석결과

3.1 어휘다발의 구조적 분류

본 연구에서는 Biber *et al.*(1999: 997)이 대화(conversation)과 학술산문(academic prose)에 나타난 어휘다발의 구조적으로 분류한 방식에 따라서 한국인과학자와 원어민과학자코퍼스에서 발견된 어휘다발을 분류하였다.

4) Biber *et al.*(2002)는 대화(conversation)장에서 lexical bundle은 여러 많은 대화 속에서 반복되는 반면 local repetitions는 특정주제와 관련된 대화에서 화자의 반복적인 사용경향이라고 하였다.

<표1> 한국인과학자와 원어민과학자 코퍼스에 나타난 어휘다발의 구조적 분류

구조	한국인과학자		원어민과학자	
	종류	비율	종류	비율
명사구조(Noun structures)	86	27%	122	35%
명사 + of phrase fragment	65	20%	94	27%
명사 + other post-modifier fragment	12	4%	14	4%
기타 명사	9	3%	14	4%
전치사구조(Prepositional structures)	31	9%	59	17%
전치사 + of phrase fragment	7	2%	14	4%
기타 전치사	24	7%	45	13%
동사구조(Verb structures)	187	58%	153	45%
수동태 + Prepositional fragment	64	20%	43	13%
기타 수동태	22	7%	20	6%
인칭대명사 we + 동사	13	4%	5	1%
계사동사 be (+ 명사/형용사)	16	5%	11	3%
(동사/형용사+) to 부정사	40	12%	47	14%
(명사+) (동사+) that 절	25	8%	13	4%
선행의(Anticipatory) it (+ 동사/형용사)	3	1%	6	2%
부사절	4	1%	8	2%
기타구조(other structures)	19	6%	10	3%
형용사구	7	2%	4	1%
기타표현	12	4%	6	2%
합계	323	100%	344	100%

분석결과를 살펴보면 전치사와 명사의 비율은 원어민과학자가 높게 나타나고 동사구의 비율은 한국인이 높은 것을 알 수 있다. 이렇게 한국인 과학자가 동사의 사용이 많은 이유는 한국어의 특성과 관련지어 생각해 볼 수 있다. 최준·송현주·남길임(2010: 175)은 한국어 학술교양서 100만 단어, 학술구어 22만 단어, 그리고 자유대화 29만 단어로 구성된 코퍼스에서 5개 형태소 연속 어휘다발을 조사하였다. 분석결과 한국어 학술문어에서 동사구 정형 표현이 81.5%를 명사구 정형 표현이 10% 그리고 문장 경계 정형 표현이 8.5%를 차지하였다. 이는 한국어의 정형 표현의 유형 중 동사구 정형 표현이 가장 높은 비율을 차지한다는 것으로 Biber *et al.*(1999)가 학술 산문에서 명사구와 전치사구 기반 어휘다발이 거의 대부분을 차지하며 동사구 기반 어휘다발이 적게 나타난 결과와 차이를 보인다. 이와 같이 동사구의 사용이 지배적인 한국어의 특성이 영어를 사용할 때에도 명사구 전치사구보다 동사구를 선호하게 하는 요인이 된 것으로 볼 수 있다.

3.2. 명사구조 (Noun structures)

명사로 이루어진 어휘다발은 학술작문에서 매우 빈번하게 발견되는 구조로서 장소, 크기, 수량 등 물리적인 묘사(the location of, the size of, the number of), 존재(the presence of, the absence of), 관계(the relationship between, the association of), 추상적인 성질(the value of, the ability of), 절차(the process of, the progression of)등 매우 다양한 의미를 나타낸다(Biber *et al.*(1999: 1015).

학술작문에서 명사구의 사용비중이 매우 크다는 사실은 이미 알려져 있는데, Biber *et al.*(1999: 996)은 학술작문에 나타난 어휘다발을 조사한 결과 명사구가 30%가량 사용되었음을 발견하였다. 이와 같이 명사의 비중이 무거운 현상은 학술작문이 동사중심 (“verb-centric”)이라기보다 명사중심 (“noun-centric”)에 가깝기 때문이다(Coxhead & Byrd 2007: 134).

본 연구에서 사용된 명사구 어휘다발의 비율 또한 예상과 같이 매우 큰 것으로 나타났으며, 한국인과학자가 86개 26%인데 비해 원어민과학자는 122개 35%로 한국인과학자가 원어민에 비해서 명사구 어휘다발의 활용도가 떨어지는 모습을 보였다. 한국인과학자의 명사구 사용이 원어민과학자에 비해서 적은 이유를 알아보기 위하여 명사구 어휘다발을 전체적으로 살펴본 결과 몇 가지 요인이 작용하고 있는 것으로 나타났다.

첫째, 선호하는 어휘다발의 과도사용(overuse) 때문으로 볼 수 있다. 먼저 한국과학자 어휘다발 가운데 원어민과 비교하여, 통계적으로 유의미할 정도로 과도사용한 the extent of 와 the effect of를 대상으로 그 이유를 찾아보았다.

<표2> 한국과학자코퍼스에서 과도사용된 명사구 어휘다발

구 분	한국과학자	원어민과학자	Log-Likelihood ⁵⁾
the extent of	156	35	44.88+
the effect of	376	191	25.19+

이를 위해 먼저 extent를 포함한 어휘다발을 검색⁶⁾하여 본 결과, the extent of 가 과도 사용된 이유는 후치수식어와 관련이 있었다. 즉 한국과학자의 경우 the extent 다음 후치수식어를 사용할 경우 of phrase만을 사용하였으나, 원어민과학자는 the extent 다음 후치수식어로 of phrase 뿐만 아니라 the extent to which, to the extent that와 같이 다양한 보충절을 사용하였다. 이것은 Chen & Baker (2010: 35)의 연구에서 원어민코퍼스에서 발견된 the extent to which, the degree to which 등의 관계절 표현을 비원어민이 사용하지 못했

5) 랑카스터대학에서 제공하는 Log-likelihood calculator를 사용하였는데, 계산 값이 6.63보다 크면 p값이 < 0.01 범위에 있게 된다. + 표시는 과도 사용을 나타낸다(출처: <http://ucrel.lancs.ac.uk/llwizard.html>).

6) 빈출기준에 따라 선별된 어휘다발 뿐만 아니라 클러스터(cluster)에서 추출한 모든 어휘다발을 대상으로 하였다. 그 이유는 결국 the extent 다음 of 이외의 수식어가 있느냐 없느냐가 the extent of의 산출빈도에 영향을 주기 때문이다.

음을 보인 연구결과와 일치한다. 따라서 한국과학자가 the extent와 결합하는 관계대명사 구문에 익숙하지 않음으로 인해 높은 사용빈도가 나타났다고 볼 수 있다.

(1) **The extent of** colocalization signal intensity is quantified in the full image as well as in region 1~3 marked in (A) to (C), respectively. (KLS, 30.txt)

(2) Content validity is **the extent to which** a measure adequately and comprehensively measures what it claims ... (Biomed, 1472-6882-3-1.txt)

(3) **To the extent that** FWER provides appreciably more stringent selection, attendant findings of significant differential gene expression ... (Biomed, 1471-2164-3-28.txt)

두 번째 the effect of의 경우는 대체 가능한 표현의 사용과 관련이 있었는데, 원어민코퍼스에서는 유사한 의미로 대체가능한 the influence of가 사용된 반면 한국인과학자코퍼스에서는 발견되지 않았고, 이로 인해 the effect of 의 사용빈도가 높아지고, 어휘다발의 종류 또한 다양하지 못한 것으로 나타났다. 이 같은 현상은 비원어민 학습자들이 자신들에게 익숙하고 확신을 가지는 표현을 과다 사용하는 의존의 섬(islands of reliability) 현상으로 설명할 수 있다(H. Dechert 1984: 227).

(4) We then monitored **the effect of** mutations on the expression of three target genes of Zur. (KLS, 35.txt)

(5) We examined **the influence of** TNF- α on monocyte migration induced by RA SCL. (Biomed, ar147.txt)

3.3. 전치사구조 (Prepositional structures)

전치사구조는 명사와 더불어 학술장르에서 가장 빈번하게 출현하는 어휘다발의 하나로서 다양한 의미적 기능을 수행 하는데 그 역할로서 추상적(as a result of, as a function of), 논리적(in terms of, in contrast to)관계, 시간(at the same time, at the end of),

문맥 지시(in this study, in the present study), 수량(in the number of, the magnitude of)등을 표현한다(Biber *et al.* 1999: 1018; Biber, Cornard & Cortes 2004: 388).

전치사구조는 한국인과학자가 31개 10%를 원어민과학자 59개 17%로 한국인과학자가 종류와 빈도에서 어휘다발의 사용이 낮은 것을 알 수 있는데, 명사와 전치사 구조는 학술작문에서 높은 비율로 나타남에도 불구하고, 이렇게 한국인의 사용비율이 낮은 이유는 무엇일까? 이것은 몇 가지 원인 때문인 것으로 밝혀졌는데, 먼저 원어민과학자에서 발견되는 대체표현을 한국인과학자들은 사용하지 않아 특정 어휘다발을 과도사용(overuse)한 것과 관련이 있었다. 두 코퍼스에서 공통적으로 사용한 어휘다발을 전체적으로 살펴본 결과, in this study의 경우 한국과학자가 276회, 원어민과학자는 235회로 41회 차이가 났는데, 원어민의 경우 대체 가능 표현인 in this report을 사용하고 있어 상대적으로 낮은 빈도를 보였다고 할 수 있다.

(6) **In this study**, we sought to identify which probiotics or mixture of probiotics could confer potent antiinflammatory effects by increasing CD4+Foxp3+ regulatory T cells (Tregs). (KLS, 5.txt)

(7) Conclusion Data presented **in this report** suggest that factors related to the sex of an individual modify the effect of smoking on the risk of RA. (Biomed, ar750.txt)

두 번째, 과도하게 사용된 어휘다발 가운데는 인칭관련 어휘가 포함된 경우가 3개 (in this study we, in our study, in addition we)나 발견되었고, 한국인과학자가 2개 이상 높은 빈도로 사용하였다.

<표3> 인칭관련 어휘다발 사용빈도 비교

구 분	한국인과학자	원어민과학자
in this study we	112	39
in our study	96	52
in addition we	56	31
합계	264	122

그 이유는 무엇일까? 학술작문에서의 인칭대명사의 사용은 능동적으로 자신들의 연구수행 및 연구결과를 주장하기 위하여 사용되는 유용한 표현방식의 하나로 인식되지만 (Hyland 2008b; 2001), 인칭대명사의 지나친 사용은 구어체의 특성(Biber *et al.* 1999: 333)에 가깝다고 볼 수 있으며, 비원어민들의 학술작문에서 사용역(register)에 관한 미숙한 사용능력 때문에 발생하는 현상으로 알려져 있다(Gilquin & Paquot 2007; Granger 1998). 이러한 관점에서 살펴보면 한국인과학자들이 비원어민 학습자로서 학술장르에 충분히 익숙하지 않아

더 많은 인칭관련 어휘다발을 사용한다고 볼 수 있을 것이다.

3.4. 동사구조(Verb structures)

3.4.1. 수동태 + 전치사구(Prepositional fragment)

자연과학과 같이 과학실험을 중심으로 하는 학술장르(academic genre)에서는 연구과정을 통해 실험자의 영향을 최소화 하면서 객관적인 결과를 보여주어야 하기 때문에 수동태구문을 많이 사용한다. 수동태는 크게 현재동사와 과거동사로 나누어 그 의미역할을 구분할 수 있는데, 먼저 수동태가 현재동사와 쓰인 경우는 표나 그림 따위의 자료를 지시(is shown in, are listed in)하거나 논리적 관계(is based on, is supported by)를 나타내는 종류가 많았고, 수동태가 과거동사와 결합하는 경우는 대부분 실험의 과정(were treated with, was performed using)을 나타내는 용도로 사용되었다.

(8) Mice with AD or IBD **were treated with** IRT5 or PBS, and the population of CD4+Foxp3+ T cells at the inflammatory sites ... (KLS, 5.txt)

(9) Finally, Southern blot analysis **was performed using** a probe designed to a Pst I restriction fragment length polymorphism ... (Biomed, gb-2003-4-8-r50.txt)

본 연구에서는 한국인과학자가 86개 27% 원어민과학자가 63개 19%를 사용하여 한국과학자가 8% 더 많이 사용한 것으로 나타났다. 이렇게 높은 비율을 차지하는 요인을 찾기 위해 어휘다발을 조사한 결과, 이는 한국인과학자의 과거형 수동태 사용과 관련이 있었다. <표4>에서 한국인과학자가 원어민보다 2배 이상 많은 어휘다발을 사용하고 있음을 알 수 있다.

<표4> 과거형 수동태(was/were + verb) 어휘다발 사용양상

구분	한국인과학자	원어민과학자
종류	45	23
빈도	2824	1202

이러한 결과로 말할 수 있는 사실은 한국과학자들은 원어민과학자에 비해 실험과정에 더욱 많은 부분을 할당하고 있다는 것이다. 즉 과학저널의 특성상 객관적인 실험을 통해 주장의 근거를 마련하기 때문에 개인의 의견이나 해석보다 실험과정에 중심을 둘 수밖에 없는데, 한국과학자들은 실험절차와 과정을 원어민과학자에 비해서 중요시하여 많이 사용한 것을 이러한 관찰을 통하여 발견할 수 있는 것이다.

3.4.2. 인칭대명사 we + 동사

인칭대명사 we + 동사가 보충절(that or whether)과 결합한 형태의 어휘다발이 원어민과 비원어민코퍼스에서 공통적으로 발견되었다. 인칭대명사 we는 학술장르에서 예상보다 많이 발견되는데 저자 자신을 언급하거나 일반화하는데 사용되며(Biber *et al.* 1999: 333), 일종의 수사적인 전략으로서 저자의 확신을 표현하거나 해당 학문분야에 자신들의 기여를 강조함으로써 그들의 권위를 주장하기 하는 도구로 사용한다고 볼 수 있다(Hyland 2008b: 554).

본 연구에서 확인된 인칭대명사 we + 동사 어휘다발은 한국인과학자가 13개 4%, 원어민과학자가 5개 1%로 종류와 빈도에서 통계적으로 유의미할 정도로 많은 사용을 보였다.

<표5> 인칭대명사 we + 동사의 사용양상

구 분	한국인과학자	원어민과학자	Log-Likelihood
종 류	12	5	118.56+
빈 도	704	235	

구체적으로 살펴보면 공통적으로 사용된 2개(we found that, we show that)를 제외하면 한국과학자는 11개, 원어민과학자는 3개의 어휘다발을 사용하였다.

<표6> we + 동사 구문의 사용종류와 빈도

구 분	한국인과학자	원어민과학자	
어휘다발	WE FOUND THAT WE HYPOTHEZIZED THAT WE PROPOSE THAT WE SHOW THAT WE TESTED WHETHER WE DEMONSTRATED THAT WE EXAMINED WHETHER WE OBSERVED THAT WE SHOWED THAT WE SUGGEST THAT HERE WE SHOW THAT WE INVESTIGATED WHETHER NEXT WE EXAMINED	WE FOUND THAT WE CONCLUDE THAT WE WERE ABLE TO WE HAVE SHOWN WE SHOW THAT	
합계	종류	13	5
	빈도	740	235

인칭대명사 we의 빈번한 사용은 한국과학자 뿐만이 아니라 비원어민 학습자에게서 나타나는 현상으로 볼 수 있다. Gilquin & Paquot(2007: 4)은 비원어민(International Corpus of Learner English)의 학술작문을 원어민과 비교한 연구에서 비원어민은 자신들이 저자임을 나타내기 위해 인칭대명사가 포함된 표현을 과도하게 사용하는데 이렇게 과도한 인칭대명사의 활용은 구어체의 특성에 가깝다고 하였으며, Granger(1998: 153)도 모국어가 불어인 비원어민과 원어민코퍼스를 비교한 연구에서 I/we/you/one + that 구조를 조사하였는데, 비원어민이 3배 이상 그리고 통계적으로 유의할 정도로 더 많은 어휘다발을 사용하였음을 밝혔

다.

그러므로 한국과학자들의 다양한 그리고 높은 빈도의 we의 사용은 비영어권 학술작문에서 나타나는 특성과 유사하며 비영어권 학습자로서 논리나 관점을 도입할 때 자신들이 사용에 확신을 가지고 있는 고정된 어구나 표현에 집착 “cling on” 하는 경향으로 볼 수 있다 (Granger 1998: 154). 정리하면, 1인칭 대명사의 사용이나 자신의 연구를 스스로 인용하는 현상은 문체적 선택의 문제만이 아니라 연구결과에 대한 인정을 획득하는 동시에 학자로서 정체성을 증진시키는 요소(Hyland 2001: 223)임에는 틀림없다. 하지만 과도하게 나타난 1인칭대명사 + we 어휘다발은 자칫 독자들에게 주관성을 내세울 수 있고, 장르차이를 반영하지 못하는 것으로 비취질 수 있기 때문에 적절한 사용이 요구된다.

3.4.3. Copula be + 형용사구(adjective phrase)/명사구(noun phrase)

형용사구와 결합하는 경우는 원인(is due to), 비교(is similar to), 명제에 관한 저자의 평가(is essential for, is important to)등의 의미로 쓰였고, 명사구가 나올 경우는 보통 of phrase와 함께 주격보어 (is one of)로 사용되었다. Copula be + 형용사(adjective) 구조는 특히 학술작문과 픽션(fiction)에서 자주 사용되며 이 가운데서도 be + 형용사와 결합하는 구조가 가장 많다(Biber *et al.* 1999: 437).

본 연구에서는 한국인과학자가 16개 5%, 원어권과학자가 11개 3%개로 한국인과학자가 5개 많은 어휘다발을 사용하고 있었다. 세부적으로 살펴보면, 두 코퍼스에서 공통적으로 사용한 어휘다발은 is consistent with와 be due to로 논리적인 관계를 설명하는 종류였다. 원어권 과학자는 평가(is important to), 헤지(is likely that), 비교(is similar to)와 같이 다양한 의미로 사용한 반면 한국인과학자는 독자적으로 사용한 14개 어휘다발 가운데 10개가 저자의 태도(attitude)나 평가(evaluation)을 나타내는 경우로 is necessary for, is critical for, is essential for, is important for, is possible that, (be/is) responsible for, an important, there were no significant)였고, 다음으로 비교(were similar to, there were no difference, significantly lower)관련 어휘다발을 많이 사용하였다.

(9) Histone H1 **is necessary for** folding chromatin into a more compact structure that becomes refractory to transcription ... (KLS, 27.txt)

(10) We examined whether the normal function of dopaminergic neurons **is essential for** normal TPB. (KLS, 32.txt)

3.4.4. (동사/형용사+) to 부정사

학술작문에서 많이 발견되는 (동사/형용사+) to 부정사 구조는 크게 3가지로 분류할 수

있다(Biber *et al.* 1999: 1023). 첫 번째, 능력이나 가능성을 의미하는 서술형용사(predicative adjective) + to 부정사로서 가능성(is likely to), 능력(were able to)을 나타낸다. 두 번째, 동사(verb phrase) + to 부정사는 수동태와 능동태에 따라 그 의미가 다르게 나타났는데, 수동태와 함께 나타날 경우 실험에서 발견한 사실(was found to be, was shown to)이나 선행연구결과(have been reported to, have been shown to)를 확인하는데 사용하였고, 능동태와 쓰일 경우 헤지 의미의 동사(appear to be, seems to be)와 결합하여 저자의 판단을 유보하는 의미로 사용되었다. 세 번째, 단순(simple) to 부정사인데, 주로 목적(to determine whether, to address this)을 나타내는데 사용되었다.

전체적인 사용에서 한국인과학자는 40개 12%를 원어민과학자는 47개 14%로 원어민과학자가 7개 더 많이 사용하였다. 그 원인을 찾아보기 위하여 to 부정사와 결합하는 형용사와 동사를 살펴보자 흥미로운 현상을 발견할 수 있었는데, 그것은 헤지 의미의 형용사와 동사의 사용에 있어서 원어민과학자가 더 많은 종류를 2.5배 이상 높은 빈도로 사용하고 있었다는 것이다.

<표7> to 부정사 어휘다발의 사용양상 비교

구 분	한국인과학자	원어민과학자	
어휘다발	is likely to(40) likely to be(68) be able to(36)	is likely to(55) likely to be(102) are likely to(34) more likely to(54) likely to have(31) be able to(34) were able to(54) were unable to(33)	
합계	종류	3	8
	빈도	144	397

이러한 결과는 비원어민 학습자들이 헤지 표현을 사용을 원어민만큼 충분히 활용하지 못했다는 선행연구 결과(Chen & Baker, 2010; Ädel & Erman, 2011)와 일치하는 것으로서 주장에 관한 책임을 약화 시키며 독자에게 판단의 여지를 남기는 기능을 헤지표현을 한국인과학자는 원어민과학자만큼 자주 사용하지 않음을 알 수 있다.

3.4.5. (명사+) (동사+) that 절

that절과 결합하는 구조는 2가지 경우로 나누어서 살펴볼 수 있다. 첫째, 명사 + that절의 경우 실험에서 발견한 사실이나 주장하고자 하는 내용을 강조하는데 사용되며 둘째, 동사 + that절이 결합하는 경우 저자의 주장 또는 선행연구결과를 추론하는데 사용하는데 과학(science)과 공학(engineering)저널에서는 높은 비중의 indicate, suggest, show 동사를 사용하여 think, believe, suspect 등의 인식 동사를 사용했을 때 보다 저자의 주장이 덜 주관적임을 나타내고 무생물(inanimate)주어와 쉽게 결합하는 특징이 있다(Hyland 1998:

366).

본 연구의 조사결과 한국인과학자는 25개 8% 원어민과학자는 13개 4%를 사용하여 숫자적인 측면에서는 한국인과학자의 사용이 많았으나, 내용적으로는 상당히 유사함을 보였다. 첫째, 명사 + that의 경우 한국과학자에서만 사용된 1개(the notion)를 제외한 모든 4개의 명사(the hypothesis, the possibility, the observation, the fact)를 공통적으로 사용하여 큰 차이를 보이지 않았다. 둘째, 동사 + that의 경우 Hyland(1998: 366)의 주장과 동일하게 두 그룹이 모두 suggest, indicate, show 3개 동사를 많이 사용하였다. 즉 한국인과학자의 70%, 원어민과학자에서는 80%가 위 세 개의 동사를 사용하였다. 다만 한국인과학자에서 어휘다발이 많이 발견된 이유는 원어민과학자에 비해서 다양한 주어가 사용된 것 때문으로 확인되었다.

3.4.6. 가주어 (Anticipatory) it (+ 동사/형용사)

가주어 it이 사용되는 구조는 2가지로 경우로 나누어 볼 수 있는데, 첫째 it + 형용사구 (adjective phrase), 동사구(verb phrase)가 와서 가능(possible), 중요성(important), 헤지(hedge) 등의 의미를 나타내며 둘째, it + 수동태 완료형(has been reported, has been shown)과 결합하여 제시된 정보의 출처를 나타내는 의미로 자주 활용된다(Biber *et al.* 1999: 1019).

과학논문에서 가주어 it 패턴(pattern)을 사용하는 이유는 저자가 직접적으로 드러나지 않으면서 명제에 관한 평가를 전달하는 기능이 있기 때문이다(Hyland 2008a: 11).

본 연구에서 한국인과학자가 3개 1% 원어민과학자가 6개 2%로 원어민과학자가 더 많은 어휘다발을 사용하였다. 한국인과학자의 경우 it + is possible(that), has been reported 처럼 형용사와 동사가 하나씩 사용된 반면, 원어민과학자의 경우 it + is possible(that), important to, likely to 형용사 3개, appears that, has been shown 동사 2개가 사용되었다. 가주어 it과 결합하는 형용사와 동사를 관찰한 결과 한국인과학자들이 헤지 형용사와 동사 각각 1개씩 사용하지 못하고 있음을 알 수 있었다. 헤지표현의 유창성과 관련하여 선행연구인 Chen & Baker(2010: 41)에서도 가주어 it + 형용사/동사 구조에서 비원어민은 원어민처럼 다양하고 왕성하게 헤지 표현을 사용하지 못했다(“as diversely and robustly as native writers do”)고 보고하였으며, Ādel & Erman (2011: 87) 또한 원어민과 비원어민을 대상으로 한 어휘다발 연구에서 원어민은 많은 수의 헤지 의미의 어휘다발을 사용할 뿐만 아니라, 다양한 종류를 사용해서 불확신 또는 의심을 표현하였다고 보고하였다. 따라서 저자가 자신의 확신이나 해석에 있어 실패할 가능성으로부터 보호할 수 있는 언어장치인 (Hyland 2008a: 18) 헤지 표현의 과소 사용은 한국인과학자들이 보다 정교한 언어사용적인 측면에서 원어민에 비해서 부족한 현상을 보여주고 있음을 보여주는 것이라 할 수 있다.

3.4.7. 부사절

부사절은 접속사 as를 사용한 어휘다발이 한국인과학자와 원어민과학자에게서 공통으로 발견되었다. 학술작문에서 as가 이끄는 부사절은 앞서 나온 표나 그래프 등을 지시하는 역할을 한다(Biber et al, 1999: 1023). 조사 결과 한국인과학자가 4개 1% 원어민과학자가 8개 2%를 사용한 것으로 나타났는데 이 가운데 4개(as described in, as described previously, as previously described, as shown in figure)가 공통적으로 사용되었다. 나머지 원어민과학자만 사용한 어휘다발 4개도 앞서 사용한 동사가 대부분 사용되었으나, 뒤에 따라오는 전치사의(in fig, by)형태가 달라서 발생하는 경우로 밝혀져서, 두 코퍼스에서 거의 동일한 역할로 사용했음을 알 수 있었다.

(11) The decrease in the current of the ZnO NW is most likely due to the negatively charged AFP antigen over the PAC shell, **as described in** Figure 3f. (KLS, 3.txt)

(12) **As shown in Figure** 1, the degree to which the nucleotide-binding site was altered is remarkable. (Biomed, gb-2003-4-3-r18.txt)

3.5. 형용사구 및 기타표현

형용사구는 한국인과학자가 7개 2%, 원어민과학자가 4개 1%로 한국인과학자가 3개 많이 사용한 것으로 조사되었다. 양쪽 코퍼스 모두에서 가장 빈번하게 나타난 어휘다발은 similar to를 포함한 종류였는데, 한국인과학자의 60%, 원어민과학자의 50%에 해당한다. 이는 실험에서 얻어지는 여러 데이터를 비교 분석해서 객관적인 자료를 제시해야하는 자연과학 장르의 특성에서 비롯된 것으로 보인다.

(13) This dramatic spectral shift **was similar to that** seen on acid denaturation of the blue-absorbing (Pb) state of blue/green CBCRs ... (KLS, 47.txt)

(14) That the effects of diamide **were similar to those** of different environmental conditions depending on the genes analyzed reflects the diverse effects of this drug on the cell ... (Biomed, gb-2002-3-11-research0059.txt)

어느 쪽에도 분류 할 수 없었던 기타 표현은 한국인과학자가 12개 4% 원어민과학자가 6개 2%였다. 종류로는 비교급(as well as, higher than that), 동사구(play a role in, play a

critical role, examined the effect of) 등의 어휘다발이었다.

4. 결 론

지금까지 코퍼스 언어학적 조사방법에 의해 한국인과학자와 원어민과학자코퍼스에서 추출한 3~4단어 어휘다발을 대상으로 출현빈도, 구문형태, 활용의미에 따라 공통점과 차이점을 살펴보고 그러한 특징들을 발생하게 하는 언어적인 요인들에 대해 설명을 시도하였다.

첫째, 한국과학자코퍼스에서 특정 어휘다발의 과도사용을 확인하였다. 명사와 전치사구조에서 특정어휘다발의 과도사용은 어휘다발의 불필요한 반복과 다양성의 부재를 가져왔다. 이로 인해 명사, 전치사구의 사용비중이 원어민과학자 및 선행연구를 통해 밝혀진 빈도수와 비교하여 상당히 낮게 사용되었음이 드러났다. 이러한 결과는 비원어민 학습자와 장르에 익숙하지 않은 원어민들이 보여준 선행연구의 결과와 유사하다(Chen & Baker 2010; Cortes 2004). 따라서 본 연구에서 원어민과학자이 사용한 명사구와 전치사구 어휘다발을 활용한 학습이 이루어진다면 보다 자연과학 학술논문에 가까운 모습을 지닐 것으로 생각한다.

둘째, 한국과학자코퍼스에서 헤지 표현의 과소사용을 확인하였다. to 부정사와 가주어 it 구조에서 함께 사용되는 헤지 형용사, 동사의 사용빈도와 종류가 원어민과학자코퍼스에 비해 부족한 모습을 나타내었다. 이것은 헤지 표현이 가지는 의미를 충분히 숙지하지 못했거나, 다른 동사나 부사를 활용하여 표현했을 가능성으로 볼 수 있을 것이다. 헤지 표현이 다른 단어나 더 짧은 단어로 표현 되었는지에 대해서는 추후 두 코퍼스의 어휘목록을 대상으로 추가 연구가 필요하리라고 본다. 셋째, 한국과학자코퍼스에서 빈도나 종류에서 더 많이 사용한 어휘다발이 발견되었다. 첫째, 인칭대명사 we + 동사의 경우로 비원어민인 한국과학자가 사용역 차이를 반영하지 못한 경우로 볼 수 있다. 둘째, 실험과정을 묘사하는데 주로 사용된 과거시제 수동태 어휘다발을 그 예로 들 수 있으며 이것은 한국과학자들이 실험과정을 더욱 많이 묘사했다는 알 수 있다. 코퍼스언어학은 이제까지 볼 수 없었던 언어사용의 실제적인 모습을 들여다 볼 수 있는 현미경과 같은 도구의 역할을 하여 언어를 관찰하고 기술함으로써 새로운 이론을 생성하거나 기존이론을 보완하는데 도움을 주는 학문이다. 앞으로 생물학분야 뿐만 아니라 여러 학문영역에 나타나는 비원어민 학습자로서 한국인이 사용한 어휘다발의 사용차이를 통해 한국인에 의해 사용되는 영어의 특성을 발견하고 그것을 바탕으로 더욱 영어다운 모습을 만드는데 도움을 줄 수 있기를 기대한다.

인 용 문 헌

- 고광윤. 2009. 코퍼스 분석을 위한 프로그램 선택: WordSmith Tools 4.0과 MonoConc Pro 2.2를 중심으로 『영어학 연구』 27, 1-22.
- 고광윤. 2005. 한국적 영어학과 영어코퍼스의 활용. 『영어학 연구』 19, 1-19.
- Ådel, A., & Erman, B. 2012. Recurrent word combinations in academic writing by native and non-native speakers of English: A lexical bundles approach. *English for Specific Purposes*, 31, 81-92.

- Biber, D., Johansson S., Leech G., Conrad, S., & Finegan E. 1999. *Longman Grammar of Spoken and Written English*. London: Longman.
- Biber, D., Conrad, S., & Cortes, V. 2004. If you look at ... : Lexical bundles in university teaching and textbooks. *Applied Linguistics* 25(3), 371-405.
- Biber, D., 2004. Historical patterns for the grammatical marking of stance. *Journal of Historical Pragmatics* 5:1, 107-136. John Benjamins Publishing Company.
- Biber, D., 2006. *University Language: A Corpus-based study of Spoken and Written Registers*. Amsterdam: Benjamins.
- Biber, D., & Barbieri, F. 2007. Lexical bundles in university spoken and written registers. *English for Specific Purposes* 26(3), 263-286.
- Cortes, V. 2004. Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for Specific Purposes*, 23, 393-423.
- Chen, Y., & Baker, P. 2010. Lexical bundles in L1 and L2 academic writing. *Language Learning and Technology* 14(2), 30-49.
- Coxhead, A. & Byrd, P. 2007. Preparing writing teachers to teach the vocabulary and grammar of academic prose. *Journal of Second Language Writing* 16, 129-147.
- De cock, S. 2004. Preferred sequences of words in NS and NNS speech. *Belgium Journal of English and Literatures(BELL)*, New Series 2, 225-246
- Dechert, H. (1984). Second language production: Six hypotheses. 211-230. In H. Dechert, D. Mohle, & M. Raupach (Eds.), *Second Language Productions*. Tübingen: Gunter Narr Verlag.
- Gilquin, G., & Paquot, M. 2007. Spoken features in learner academic writing: Identification, explanation and solution. *Proceedings of the Fourth Corpus Linguistics Conference*. 1-12. Birmingham, United Kingdom.
- Granger, S. 1998. Prefabricated pattern in advanced EFL writing: Collocations and formulae. 145-160. In A. Cowie (Ed.), *Phraseology: Theory, Analysis, and Applications*. Oxford: Oxford University Press.
- Hyland, K. 2001. Humble servants of the discipline? Self-mention in research articles *English for Specific Purposes* 20, 207-226.
- Hyland, K. 2008a. As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes* 27, 4-21.
- Hyland, K. 2008b. Genre and academic writing in the disciplines. *Language Teaching* 41(4), 543-562.
- Kennedy, G. 1998. *An Introduction to Corpus Linguistics*. London: Longman.
- Nattinger, J. R., & DeCarrico, J. S. 1992. *Lexical Phrases and Language Teaching*. Oxford: Oxford University Press.
- Renouf, A. and Sinclair, J. 1991. Collocational Frameworks in English, in K. Ajimer and B. Altenberg(eds) *English Corpus Linguistics*. Cambridge: Cambridge university Press, 128-143.

- Scott, M., 2008. *WordSmith Tools version 5*, Liverpool: Lexical Analysis Software.
- Sinclair, J. 1991. *Corpus. Concordance, Collocation*. Oxford: Oxford University Press.

발표자: 이성민

소속기관: 부경대학교

연락처: roy7942@hanmail.net