

정 소 우 교수지도
석사학위 청구논문

코퍼스 기반 원어민 화자와 한국인 화자의
학술 논문에 나타난 언어적 비교 분석

- 영어 관련 학술 논문 중심으로 -

2013

성신여자대학교 교육대학원
교육학과 영어교육전공
송 경 진

코퍼스 기반 원어민 화자와 한국인 화자의
학술 논문에 나타난 언어적 비교 분석

- 영어 관련 학술 논문 중심으로 -

정 소 우 교수지도

이 논문을 석사학위논문으로 제출함

2012년 11월

성신여자대학교 교육대학원
교육학과 영어교육전공
송 경 진

인 준 서

송경진의 석사학위 논문으로 인준함.

심사위원 _____(인)

심사위원 _____(인)

심사위원 _____(인)

성신여자대학교 교육대학원

논문 개요

외국어로써 영어를 배우는 환경에서 원어민처럼 말하고 글을 쓴다는 것은 결코 쉽지 않은 목표일 것이다. 특히 교육과 학문이 발달하고 다양해짐에 따라 학문적 글쓰기가 많이 이루어지고 있는 요즘 자신의 연구 결과와 생각을 논리적으로 표현해내는 것이 상당히 중요하다고 할 수 있다. 따라서 본 논문은 코퍼스를 기반으로 하여 원어민 화자와 한국인 화자의 학술 논문에 나타난 다양한 언어적 특징들을 비교하였다. 연구의 목적은 이러한 특징들의 비교 결과를 통해 학문적 글쓰기를 하는데 있어서 원어민과 한국인 사이의 차이를 파악하고 보완하며 영어 교육의 측면에서도 이를 적용할 수 있도록 하는데 있다. 이를 위해 다양한 학문 영역 중 영어 관련 학문을 선택하여 각각 영어영문학, 영어학, 영어교육학 분야의 원어민 화자와 한국인 화자의 학술 논문을 수집하여 분석하였다. 그 결과는 다음과 같다.

첫째, 어휘의 측면에서 비교해 본 결과, 내용어의 경우 명사는 비슷한 어휘가 많이 확인되었지만 원어민 화자의 학술 논문에서 좀 더 다양한 동사와 형용사의 사용이 확인되었다.

둘째, 기능어 중에서도 대명사만 따로 확인해 보았는데 인칭대명사로는 원어민 화자의 학술 논문에서 'we'가 한국인 화자의 학술 논문에 비해 두 배 이상 나타났고 지시대명사 'these'와 'those'의 경우 원어민 화자가 한국인 화자에 비해 'these'를 더 많이 사용하는 것으로 나타났다.

셋째, 의사소통에 있어 상당히 중요한 역할을 하는 서법조동사에 대해 알아보았는데 전체적으로 원어민의 학술 논문에서 서법조동사가 좀 더 사용되었고 세부적으로는 원어민 화자의 경우 조동사의 과거형인 'could, would, might'의 사용이 한국인 화자에 비해 훨씬 많았고 의무를 나타내는 'should'와 'must'에 있어서는 원어민 화자에 비해 한국인 화자가 'should'의 사용이

훨씬 많이 나타났다,

넷째, 어휘의 측면에서 비교해 본 결과, 내용어의 경우 명사는 비슷한 어휘가 많이 확인되었지만 원어민 화자의 학술 논문에서 좀 더 다양한 동사와 형용사의 사용이 확인되었다.

다섯째, 원어민과 같이 자연스럽게 유창한 영어를 위해서 반드시 필요한 어휘 묶음에 대해 연구하였는데 그 사용 빈도, 구조, 기능의 측면에서 비교해 보았다. 먼저 사용 빈도면에서 학문별로 원어민 화자의 학술 논문의 경우 영어학에서 훨씬 많은 어휘 묶음이 나타났고 한국인 화자의 학술 논문의 경우 영어교육학에서 훨씬 많은 어휘 묶음이 나타났다. 원어민의 경우 'at the same time'이 가장 많은 빈도수로 나타났고 한국인의 경우 'on the other hand'가 압도적으로 많이 나타났다. 다음으로 구조의 측면에서 보면 두 코퍼스 모두 전치사 'of'가 포함된 명사구와 전치사구의 구조가 가장 많았고 다른 전치사가 사용된 구역시도 공통적으로 많이 나타났다. 마지막으로 기능의 측면에서 보면 두 코퍼스 모두 전체적으로는 지시 어휘 묶음이 가장 많이 나타났다. 하위 범주에서는 원어민 화자의 경우 한국인 화자에 비해 양화 어휘 묶음을 더 많이 사용한 것으로 나타났다. 그리고 텍스트 구성 어휘 묶음의 하위 범주인 틀형성 기능의 어휘 묶음이 두 코퍼스에서 모두 가장 많이 나타났다.

여섯째, 결속력 있는 글을 쓰기 위해 반드시 필요한 접속(부)사에 대해 살펴보았다. 'however'의 경우 빈도수에는 큰 차이가 없었으나 사용 위치에 있어서 원어민의 경우 문두와 문중의 사용이 비슷한 반면 한국인의 경우 문두의 사용이 훨씬 많았다. 'so'의 경우 비슷한 빈도로 나타났고 사용 위치는 문중이 많았다. 그리고 'therefore'과 'thus'의 경우 역시 빈도수에는 큰 차이가 없었으나 'therefore'의 위치의 경우 원어민은 문중, 한국인은 문두의 경우가 압도적으로 많았다.

일곱째, 참조 코퍼스를 이용해 중요 어휘를 살펴보았는데 BNC를 참조한 결과 공통적으로는 ‘student, teacher, learner, as, of, in’이 나타났고 원어민의 경우 ‘dialect, variation, texts, frequency’등의 어휘가, 한국인의 경우 ‘feedback, proficiency, native, syntactic, constructions, clauses, adjunct’등의 어휘가 각각 나타났다. 또한 각각을 서로의 코퍼스로 이용한 결과 원어민의 경우 ‘we, these, our, would’등이, 한국인의 경우 ‘the, Korean, feedback, EFL, proficiency, his’등이 나타났다.

본 연구 결과를 토대로 본 논문에서는 원어민과 한국인 사이의 영어 사용에 있어서 차이를 극복하기 위해 학습자에게 다양한 장르를 통한 진정성 있는 영어를 접할 수 있는 기회를 많이 마련해 주어야 한다고 제시하였고 교수자 또한 이러한 차이점을 인식하고 제대로 가르친다면 영어 학습자 입장에서 결코 쉽지 않은 학문적 글쓰기, 더 나아가 다양한 장르의 글을 쓰는데 있어서도 도움이 될 것이라고 하였다.

목 차

논문개요

I. 서론	1
1. 연구의 필요성 및 목적	1
2. 연구 문제	2
II. 이론적 배경 및 선행연구	4
1. 이론적 배경	4
1.1 코퍼스 언어학	4
1.2 주요 언어 코퍼스	5
2. 선행 연구	6
2.1 코퍼스 관련 연구	6
2.2 원어민과 비원어민 간의 비교 연구	8
2.3 어휘 묶음에 관한 연구	11
2.4 코퍼스 관련 수업 활용 연구	15
III. 연구 방법 및 실제	18
3.1 연구 대상 및 자료수집	18
3.2 연구 도구	20
3.3 연구 절차와 방법	21
3.3.1 어휘 비교	21
3.3.2 어휘 묶음 비교	22
3.3.3 중요 어휘 비교	22
IV. 연구 결과 및 논의	24
4.1 어휘 목록과 기능어 사용의 비교 분석	24
4.1.1 어휘 목록 비교 분석	24
4.1.2 기능어 사용의 비교 분석	27

4.2 대명사와 서법조동사 사용의 비교 분석	29
4.2.1 대명사 사용의 비교 분석	29
4.2.2 서법 조동사 사용의 비교 분석	31
4.3 어휘 묶음 사용의 비교 분석	34
4.3.1 어휘 묶음의 사용 빈도	35
4.3.2 어휘 묶음의 구조에 따른 사용 비교 분석	51
4.3.3 어휘 묶음의 기능에 따른 사용 비교 분석	54
4.4 접속(부)사 사용의 비교 분석	61
4.5 참조 코퍼스에 따른 중요 어휘 비교 분석	66
4.5.1 BNC 참조 중요 어휘 비교 분석	66
4.5.2 원어민 화자와 한국인 화자의 코퍼스 참조 중요 어휘 비교 분석 ..	69
V. 결론 및 제언	73

참고문헌

ABSTRACT

표 · 그림 목 차

<표 1> 학술 논문 학술지 출처와 논문 수	19
<표 2> 학술 논문의 영어 분야 당 논문 개수	19
<표 3> 두 코퍼스 간의 어휘 목록 비교	25
<표 4> 기능어의 사용 빈도수 & 비율	27
<표 5> 대명사의 사용 빈도수 & 비율	29
<표 6> 서법조동사의 사용 빈도수 & 비율	32
<표 7> 학문별 어휘 유형(type) & 구현(token) 수	35
<표 8> English-Native의 학문별 어휘 묶음 빈도 & 텍스트 수	36
<표 9> English-Non-native의 학문별 어휘 묶음 빈도 & 텍스트 수	39
<표 10> 두 코퍼스 간 어휘 묶음 상위 빈도 20개	43
<표 11> 영문학 분야에서 많이 나타난 어휘 묶음	44
<표 12> 영어학 분야에서 많이 나타난 어휘 묶음	46
<표 13> 영어 교육학 분야에서 많이 나타난 어휘 묶음	49
<표 14> English-Native의 어휘 묶음의 구조에 따른 분류	51
<표 15> English-Non-native의 어휘 묶음의 구조에 따른 분류	52
<표 16> English-Native의 어휘 묶음의 기능에 따른 분류	55
<표 17> English-Non-native의 어휘 묶음의 기능에 따른 분류	57
<표 18> 접속(부)사의 빈도수 & 위치	62
<표 19> BNC 참조 두 코퍼스 간의 중요 어휘 목록 비교	66
<표 20> EN과 KN의 코퍼스 참조 중요 어휘 목록 비교	70
<그림 1> concordance plot을 이용한 ‘tarzan’의 분포	69

I. 서론

1. 연구의 필요성 및 목적

오늘날 영어의 사용은 많은 분야에 있어 너무나 중요한 위치를 차지하고 있다. 특히 우리나라의 경우는 영어를 ESL(English as a Second Language) 환경에서 공용어로 사용하는 것이 아닌 EFL(English as a Foreign Language) 환경에서 외국어으로써 학습하다보니 영어를 사용함에 있어서 모국어로 사용하는 원어민과의 다양한 차이들이 발생할 수밖에 없는 것이 현실이다. 그만큼 우리나라 사람들에게 있어서는 영어를 학습하고 익히면서 최대한 원어민에 가깝게 말하고 쓸 수 있도록 하는 것이 최대의 과제가 되고 있는 것이다. 따라서 원어민 화자와 한국인 화자의 영어 사용에 있어서 어떠한 차이점이 있는지 비교하는 연구는 무엇보다 이러한 우리의 목적을 위해서는 반드시 필요하다고 볼 수 있다. 여러 방면에 걸쳐 원어민과 비원어민의 영어 사용을 비교하는 연구들이 많이 나타나고 있는 것에서도 확인할 수 있을 것이다. 또한 최근 들어 코퍼스 언어학이 발달함에 따라서 다양한 프로그램을 이용하여 대량의 언어자료를 연구할 수 있게 되면서 더욱더 활발히 이러한 연구들이 진행되고 있다.

나날이 인터넷이 발달하고 전 세계를 무대로 삼는 일이 너무도 많기 때문에 세계 공용어으로써 다양한 정보를 공유하고 사람들을 만나야 함에 있어서 꼭 필요한 영어를 우리는 일찍부터 학습하고 이를 바탕으로 자신이 원하는 분야에서 지식을 습득하고 다양한 연구를 하고 있다. 그러면서 자신의 연구를 많은 이들에게 알리고 공유하기 위해 많은 사람들이 다양한 학술 논문을 쓰고 있다. 특히 영어가 세계 공용어이다 보니 좀 더 많은 사람들과 이를 공유할 수 있도록 영어로 쓰는 경우가 많아지고 있는 만큼 우리가 얼마나 제

대로 영어를 사용하고 있는지 확인하는 것이 중요하다. 이런 면에 있어서 원어민의 글쓰기와 비교하여 한국인 화자의 글에서 나타나는 여러 특징들을 살펴볼 필요가 있는 것은 분명하다. 특히 학술 논문의 경우는 자신의 연구에 대한 내용을 명확하고 조리 있게 전달해야 하며 개인적으로 편안하고 자유롭게 쓰는 글쓰기가 아닌 특정 목적을 갖고 다수가 읽을 수 있도록 하는 공식적인 글이기 때문에 더욱더 정확한 어휘, 표현, 문법 등이 요구되어 진다. 따라서 영어를 모국어나 공용어가 아닌 외국어로서 학습을 해온 우리로써는 이러한 많은 것들이 요구되어지는 학술 논문을 쓰는 것이 결코 쉽지 않다. 그렇기 때문에 원어민과 한국인 화자의 학술 논문 비교를 통해 우리는 좀더 정확하고 자연스러운 학문적 글쓰기를 할 수 있을 것이다.

이러한 측면에서 원어민 화자와 한국인 화자의 학술 논문을 비교해 보는 것은 의미가 있다고 생각한다. 많은 분야가 있지만 본 논문에서는 영어를 가장 많이 학습하고 연구했다고 할 수 있는 영어 분야의 원어민 화자와 한국인 화자의 학술 논문의 코퍼스를 통한 비교, 분석을 해 봄으로써 어떠한 차이가 있는지 확인하고 이를 바탕으로 나아가 앞으로의 한국인 화자들의 학문적 글쓰기에 있어 참고와 도움이 되고자 한다.

2. 연구 문제

영어 관련 분야에 있어서 원어민 화자와 한국인 화자의 학술 논문 비교 분석을 위하여 본 논문에서 살펴볼 문제는 다음과 같다.

첫째, 원어민 화자와 한국인 화자의 영어 관련 학술 논문에 나타난 어휘들과 기능어들 간의 사용 빈도에 어떤 차이가 나타나는가?

둘째, 원어민 화자와 한국인 화자의 영어 관련 학술 논문에 나타난 대명사, 서법조동사의 사용 빈도에 어떤 차이가 나타나는가?

셋째, 원어민 화자와 한국인 화자의 영어 관련 학술 논문에 나타난 어휘 묶음(lexical bundle)의 사용과 빈도에 어떤 차이가 나타나는가?

넷째, 원어민 화자와 한국인 화자의 영어 관련 학술 논문에 나타난 접속사, 접속부사의 사용 빈도와 위치에 어떤 차이가 나타나는가?

다섯째, 원어민 화자와 한국인 화자의 영어 관련 학술 논문에 나타난 중요 어휘들에는 어떤 차이가 나타나는가?

Ⅱ. 이론적 배경 및 선행연구

1. 이론적 배경

1.1 코퍼스 언어학

코퍼스 언어학은 '실제 언어' 혹은 실제 언어의 샘플을 이용하여 언어를 공부하는 응용언어학의 한 분야이다. 말뭉치(코퍼스)란, 언어를 연구하는 각 분야에서 필요로 하는 연구 재료로서 언어의 본질적인 모습을 총체적으로 드러내 보여줄 수 있는 자료의 집합을 뜻한다. 조건만 만족할 수 있으면 작게는 시집 한 권이나 소설 한 편으로부터 1억 어절 이상의 말 또는 글로 표현된 각종의 자료에 이르기까지, 다양한 크기의 자료 모음이 모두 코퍼스라는 이름으로 묶일 수 있으며, 그 내용도 연구의 목적에 따라 다양하게 구성될 수 있다. 그러므로 코퍼스라는 용어는 연구의 목적이나 성격에 따라 다양한 대상을 지시할 수 있는 포괄적인 개념이 있다. 최근의 언어 연구에서는 코퍼스의 개념을 더 좁게 보아, "일정 규모 이상의 크기를 갖추고 내용적으로 다양성과 균형성이 확보된 자료의 집합체"를 가리키는 개념으로 사용하는 것이 일반적이다. 초기에는 수작업으로 이루어졌으나 컴퓨터의 발달로 지금은 많이 자동화되었으며, 전산언어학의 도움을 많이 받고 있다.

츨스키의 관점에서는, 실제 언어는 언어수행상의 오류로 가득하기 때문에 언어학을 제대로 연구하기 위해서는 잘 제어된 환경에서 이루어진 언어를 다룰 필요가 있다고 하였는데, 코퍼스 언어학은 이러한 관점에 정면으로 반하고 있는 방법론이다. 츨스키가 다루는 정제된 환경에서는 언어에서 화자에 대한 정보를 아무 것도 찾아낼 수 없다는 것이다. 코퍼스 언어학에서는 언어

능력(competence)과 언어수행(performance)으로 나누는 촘스키식 이분법도 거부한다. 대신 말뭉치에 다가갈 수 있는 최소한의 인터페이스만 있으면 실제 언어를 통해 언어 현상의 본질에 다가갈 수 있다고 본다.

1.2 주요 언어 코퍼스

코퍼스에는 다양한 종류가 많이 있는데 그 중에서도 현대적 의미의 본격적인 전자 코퍼스는 Brown 코퍼스로 1961년에 미국에서 출판된 미국 문어체 영어 텍스트를 토대로 미국의 Brown대학에서 구축하였으며 총 100만개의 단어로 구성되어 있다. 이후 많은 다른 코퍼스 구축의 모델이 되었는데 특히 1978년에 완성된 LOB(Lancaster-Oslo-Bergen)는 영국 영어 코퍼스로써 역시 1961년에 출판된 15가지 종류의 문어 텍스트 500개로 구성되어 있다. 또한 주목할 만한 코퍼스로는 ICE(The International Corpus of English)인데 이는 1990년 영국 런던대학의 영문과를 중심으로 만들어지기 시작하여 영어의 주요 국가 방언에 대한 전자 코퍼스 구축 작업을 진행하고 있다. 구어체와 문어체의 영어자료를 바탕으로 하는 100만개의 단어로 구성되어 있다. 다음으로 영국의 Birmingham 대학과 출판사 HarperCollins가 공동 작업으로 구축하는 Bank of English는 현재 약 5억 6천만 단어에 이르는 현존하는 영어 코퍼스 가운데 가장 방대하고 광범위하다고 할 수 있다. 또한 Oxford대학 출판부 주도아래 여러 기관의 공동연구와 작업으로 이루어진 영국영어의 구어체와 문어체 코퍼스로써 현재 가장 널리 사용되고 있는 약 1억 단어로 구성된 BNC(British National Corpus)가 있다. 이렇듯 그동안 영국 중심으로 코퍼스의 구축이 많이 이루어져왔는데 그러한 지역적 불균형을 깨면서 BNC에 상응하는 미국영어 코퍼스인 ANC(America National Corpus)가 있다. (고광윤, 2005)

2. 선행 연구

본 연구에 앞서 다양한 선행 연구를 살펴보려 한다. 선행 연구는 본 논문과 관계있는 코퍼스 관련 연구, 원어민과 비원어민간의 비교 연구, 어휘 묶음에 관한 연구, 코퍼스 관련 수업 활용 연구로 나누어서 살펴보겠다.

2.1 코퍼스 관련 연구

최근 들어 코퍼스 언어학이 발달하면서 코퍼스를 이용한 다양한 연구들이 많이 진행되고 있다. 이런 상황을 반영하여 이은주(2007)는 ‘영어교육과 응용언어학 분야에서 수행된 코퍼스 기반 연구의 분석’에서 최근 코퍼스를 기반으로 한 연구가 응용언어학에서 외국어 교육 분야에까지 적용 가능성이 지속적으로 확대되고 있는 가운데 이러한 흐름에 발맞추어 국내외에서 이루어지고 있는 코퍼스 연구의 내용과 주제 및 연구방법 등을 분류하여 코퍼스 연구의 현황을 파악하고 이를 토대로 이론적, 교육적 시사점을 제공한다. 연구의 내용을 살펴보면 우선 조사대상으로 국내외 주요 학술지에 게재된 영어교육과 응용언어학 관련 연구들 중 코퍼스를 주제로 한 연구 논문을 수집하였다. 분석 방법으로는 우선 연구 논문을 크게 연구주제, 연구 방법론, 코퍼스의 특징의 세 가지로 분류하였다. 각각의 분류대로 행한 연구 결과로 우선 연구주제 면에 있어서는 국내 코퍼스 연구 주제에는 어휘 연구가 가장 많은 비율을 차지하였고 세부적으로는 품사의 빈도수와 분포 조사, 조동사, 접속부사 등의 개별어휘 분석 등이 많았고 최근에는 연어의 분석도 많이 나타나고 있다. 또한 국제 코퍼스 연구 주제도 비슷한 양상을 보인다. 연구에서 주목할 점은 국내 코퍼스 연구의 경우 실제 현장에서의 영어교육에 코퍼스와 키크던스 프로그램 활용을 시도한 교육관련 연구가 국제 코퍼스 연구

의 경우보다 많이 나타난다는 점이다. 하지만 국내외 모두 어휘에 대한 분석이 주를 이루고 있는 만큼 어휘 수준이 아닌 폭넓은 문맥인 담화의 수준에서 코퍼스의 활용과 연구 방안이 모색되어야 한다고 제시한다. 다음으로 연구 방법론 측면에서는 국내외 모두 정량적 방법에서도 빈도와 분포를 분석한 기술통계가 주를 이루고 있다. 마지막으로 코퍼스의 특징으로 보면 국내 코퍼스는 대다수가 문어 코퍼스를 사용하였고 이는 국제 코퍼스도 마찬가지이나 편중 현상은 다소 완화된 것을 볼 수 있다. 또한 국내외 모두 학습자보다는 원어민의 코퍼스를 많이 활용하였다. 결과를 토대로 코퍼스를 활용한 연구가 두 가지 측면에서 영어교육에 기여할 수 있다고 말한다. 첫째는 영어교육 연구자가 교재를 개발하거나 영어 교사가 기존 교재를 응용하는데 적극적으로 활용할 수 있도록 촉진하는 것과 둘째, 학습자가 코퍼스를 활용하여 자기 주도적 언어학습, 즉 발견적 언어학습을 촉진시킬 수 있다는 것이다.

또한 김준기(2007)는 ‘세계 영어의 지역적 차이에 대한 말뭉치 분석 연구’에서 개별 영어의 특성을 연구하는데 있어 가장 중요한 자료 중의 하나가 말뭉치(Corpus)라고 말하며 말뭉치 언어학(Corpus Linguistics)을 이용한 연구의 가장 큰 특징은 실제 용례에 기초한 연구라는 점에서 영어 학습자에게 큰 도움이 된다고 하였다. 이 연구에서는 신문을 코퍼스로 이용하였는데 각각 한국, 미국, 영국, 호주의 신문을 수집하여 코퍼스를 구성하였다. 분석의 주는 서법 조동사의 사용이다. 분석 결과 서로 다른 지역 뿐 아니라 같은 지역의 통용어 사이에도 격식성과 같은 특성에 의해 다양성이 존재하고 있음을 입증했다. ‘must’와 ‘should’의 경우 EFL 환경에서 많이 나타나고 있고 ‘have to’는 ENL(English as a Native Language) 환경에서 빈도수가 높게 나타난다. 의미상으로는 의무의 의미를 대표적으로 나타내는 ‘must’보다 ‘should’의 사용이 더 많았고 의미상으로는 두 환경 모두 분포율이 크게

다르지 않았다. 결과를 통해 학교 문법 학습에 의해 알고 있는 서법 조동사의 강조된 의미와 실제 사용된 의미에는 차이가 있음을 알 수 있었다. 따라서 정확한 의미의 습득을 위해서는 실제 사용되는 용례들의 정확한 분석을 바탕으로 이루어져야 할 것을 주장했다.

2.2 원어민과 비원어민 간의 비교 연구

외국어로써 영어를 학습하는 사용자에게 있어서 최대한 원어민에 가깝게 영어를 구사하는 것이 목표가 될 만큼 원어민과 비원어민 간의 영어 사용에 있어서의 비교 연구는 참으로 중요하다고 할 수 있다. 그런 만큼 다양하고 많은 연구들이 이루어지고 있는데 고수원(2010)은 ‘한국인 학자와 영어 원어민 학자의 논문 영문 초록 비교 분석’에서 초록은 요약과 같이 어떤 논문을 신속하게 알 수 있도록 해주며 따라서 독자들은 초록을 기초로 하여 전체 논문을 읽을지에 대한 결정을 하는 만큼 초록의 역할이 중요하다고 말하며 연구 이유를 밝히고 있다. 따라서 각각의 초록에 나타나는 수사구조와 특징을 밝힘으로써 한국의 학자들이 효과적인 논문 초록을 작성하도록 도움을 주는데 그 목적이 있다. 연구 과제로는 수사구조와 응집성을 위한 접속사의 사용 비교 분석이다. 초록 수집은 영어교육분야로 제한하였고 영어 원어민 학자, 미국에서 수학한 학자, 한국에서 수학한 학자, 이렇게 셋으로 분류하여 각각 30개씩 총 90개의 논문 초록을 비교 분석하였다. 먼저 수사구조는 연구 배경, 목적, 방법, 결과, 논의로 분류하였는데 연구 결과 한국인 학자는 연구 배경이 빠진 수사구조를 많이 나타내고 있는 반면 원어민 학자는 모든 내용이 들어간 경우가 많이 나타나고 있었다. 다른 사항은 비슷하나 연구 배경의 유무가 가장 큰 차이를 보인다고 하겠다. 이는 미국에서 수학한 학자에서도 비슷하게 나타난다. 역시 연구 배경에 대한 언급이 많이 나타나는 것을 볼

수 있다. 다음으로 접속사 사용은 부가, 반의, 인과, 시간 접속사로 나누어 살펴보았는데 가장 큰 차이로는 한국인 학자는 응집성을 위한 접속사의 사용 중에서 부가 접속사가 반 이상을 차지하는 반면 원어민 학자는 골고루 분포하는 것을 확인하였다. 총 접속사 사용 빈도의 경우도 원어민 학자보다 한국인 학자가 더 많은 것으로 나타났다. 이 연구를 통해 외국어로서의 영어 작문 지도에 있어 단지 번역 연습에 그치지 않고 해당 국가의 논리 전개 구조도 함께 가르쳐야 하고 접속사 사용에 대한 올바른 이해와 교육이 필요하다고 제안하고 있다. 또한 감경아(2007)는 ‘원어민과 비원어민의 논문 초록에 나타난 축약 관계사절 연구’에서 관계사절의 축약현상은 많이 연구되어 왔는데 그동안의 연구는 구조적 관점에 초점을 맞췄다면 본 연구는 그 쓰임에 초점을 맞추어 실제 텍스트에서의 사용이 어떻게 나타나는지 살펴보았다. 이를 위하여 인문사회와 자연공학 분야에서의 국내외 학위 논문 초록을 수집하여 원어민은 102,372단어, 비원어민은 102,773단어로 이루어진 코퍼스를 구축하였다. 연구내용으로는 통사적인 환경이 관계사절의 축약에 어떠한 영향을 미치는지, 원어민과 비원어민 간에는 어떤 차이가 있는지, 학문 분야에 따른 차이점은 있는지의 여부이다. 먼저 관계사절의 빈도를 살펴보면 국외 학위논문 초록에서 1862개, 국내에서 1304개로 국외 데이터에서 더 많은 관계사절이 사용되었음을 볼 수 있다. 그 중에서도 주격 관계사절의 빈도수가 높는데 관계대명사와 뒤에 나오는 be동사를 생략하는 주격 관계사절의 축약은 국외 59%, 국내 48%로 원어민에 비해 비원어민의 축약 빈도가 더 낮은 것으로 확인되었다. 또한 분야 간의 사용 빈도를 확인한 결과 관계사절의 사용 비율은 자연공학 분야가 인문사회 분야보다 낮지만 축약의 경우는 61%로 가장 많이 일어나는 것을 볼 수 있다. 이는 자연공학 분야의 글이 인문사회보다 간결한 경향을 반영하는 것이라고 볼 수도 있겠다. 결과적으로 논문이나 논문초록과 같은 학문적인 글에서 비형식적인 스타일인 축약은 피해야

한다는 견해와는 달리 국내의 초록을 분석한 결과 원어민 데이터에서 축약이 많이 나타나는 것을 볼 수 있었다. 따라서 학문적 글쓰기를 가르치는데 있어 본 논문을 활용할 수 있다고 본다.

구선화(2012)는 ‘영어 원어민과 한국인 영어 학습자의 학술지 논문 코퍼스에 나타난 결과 접속 부사 사용 비교’에서 학문적인 글에서의 결과 접속 부사가 어떻게 사용되었는지를 분석하였다. 연구를 위해 한국인 학습자 50,000 단어, 원어민 학습자 50,000단어의 코퍼스를 구축하였다. 코퍼스 프로그램으로는 WordSmith Tools를 사용하였다. 연구 결과 첫째로 10개의 결과 접속 부사의 총 빈도수는 둘 다 비슷하게 나타났다. 또한 둘 다 ‘therefore’과 ‘thus’를 다른 접속 부사에 비해 많이 사용하였다. 둘째로 한국인 학습자가 원어민에 비해 ‘so’를 빈번히 사용하는 것을 볼 수 있었다. 셋째로 한국인 학습자는 결과 접속 부사를 문두에 사용하는 경우가 원어민 학습자의 경우보다 많은 것을 확인할 수 있었다. 본 연구 결과를 토대로 시사하는 바는 한국인 영어 학습자의 경우 자연스럽게 많은 인공적인 영어 환경에 처하게 되는 경우가 많은데 따라서 진정성 있는 영어를 접하기 위해서 많은 종류의 언어 자료에 노출되어야 한다. 이러한 결과 접속 부사의 사용 연구는 추후 영어 교재나 교수 요목등에 지침이 될 수 있다는 측면에서 유용한 정보를 제공한다.

아직 논문으로 발표되지는 않았으나 장세은과 홍신철(2012)은 코퍼스를 기반으로 하여 원어민과 한국인 저자의 영어학과 영문학의 학술 논문 영문 초록을 비교 분석하였다. 연구 내용을 보면 한국인 저자의 영어학과 영문학 초록 간의 언어적 특징이 무엇인지와 원어민과 비원어민 간의 어휘의 빈도, 구조, 어휘 묶음을 비교 분석하는 것이다. 연구 목적은 학술적인 영어 작문에 있어서 적절한 방법들을 제시하는데 있다. 연구를 위해 영어학과 영문학 초록을 수집하고 각각 원어민과 비원어민으로 하위분류 하였다. 어휘 목록과

중요 어휘 목록을 통해 나타난 어휘의 차이를 살피고 서법 조동사와 대명사의 빈도수를 비교하고 구조적인 범주로써 어휘 묶음의 빈도를 확인하고 많이 사용된 ‘on the other hand’, ‘as one of the’, ‘in terms of the’ 등의 빈도를 확인했다. 또한 원인과 결과를 나타내는 비슷한 표현들의 빈도도 알아 보았다.

2.3 어휘 묶음에 관한 연구

외국어로써 영어를 학습하여 말하고 쓰는 것에 있어서 일련의 연속된 단어들의 결합체인 어휘 묶음의 역할은 상당히 중요하다. 이러한 어휘 묶음을 많이 아는 만큼 좀 더 자연스럽게 유창하게 영어를 사용할 수 있을 것이다. 따라서 이에 관한 연구도 상당히 이루어지고 있는 가운데 Yu-Hua Chen과 Paul Baker(2010)는 ‘Lexical Bundles in L1 and L2 Academic Writing’에서 L1과 L2의 학문적 작문 코퍼스와 출판된 학문적 텍스트를 비교하여 어휘 묶음을 연구하였다. The Freiburg-Lancaster-Oslo/Bergen(FLOB) 코퍼스와 The British Academic Written English(BAWE) 코퍼스가 연구를 위해 사용되었다. 우선적으로 Academic Corpora를 native expert writing, native peer writing, learner writing(Chinese students)의 세 가지로 구분하였다. 프로그램으로는 Wordsmith 4.0을 이용해 4개의 단어로 이루어진 말뭉치를 찾아내었다. 나타난 결과들을 구조와 담화 기능으로 나누어 확인하였는데 우선 구조의 경우 The Longman Grammar of Spoken and Written English(LSWE) 코퍼스를 사용하였고 담화 기능은 ‘in the context of’, ‘at the same time’ 와 같은 지시 어휘 묶음(referential expressions)와 ‘are more likely to’, ‘it is necessary to’와 같은 견지 어휘 묶음(stance bundles), ‘can be used to’, ‘one of the most’ 와 같은

담화 구성 어휘 묶음(discourse organizers)로 나누어 정리하였다. 연구 결과 원어민과 비원어민 학습자 사이에 어휘 묶음의 사용이 상당히 유사하다는 것이 확인되었다. 둘 다 native expert에 비해 VP-based bundles와 담화 구성 어휘 묶음의 사용이 많았다. 이는 아직 완전치 않은 작문이라는 것을 나타내는 특징들이라고 할 수 있다. 이런 결과를 바탕으로 학문적 글쓰기에 있어서 학습자들에게 native expert의 작문에서 나타난 표현들을 학습시킨다면 좀 더 원어민적인 스타일로 글을 쓰는데 도움이 될 것으로 보인다고 말하고 있다.

또 Viviana Cortes(2004)는 'Lexical bundles in published and student disciplinary writing: Examples from history and biology'에서 우선적으로 학술지에 게재된 역사와 생물학 분야의 논문에서 많이 사용되는 4단어로 이루어진 어휘 묶음을 구조적, 기능적으로 분류, 확인하였고 다음으로 수준별 세 단계로 나눈 학생들의 작문에서 얼마나 확인된 어휘 묶음을 사용하고 있는지를 알아보았다. 연구를 위해 우선 역사와 생물학 학술지에 게재된 논문을 수집하여 코퍼스를 구축하고 다음으로 수준별로 나누어 학생들의 작문을 수집하여 또 다른 코퍼스를 구축하였다. 연구를 위해 The Lexical Bundles program(LBP)와 MonoConcPro가 사용되었다. 연구 결과 학생들이 전문적인 논문에 사용된 표현들을 많이 사용하지 않는 것으로 확인되었다. 따라서 이 둘 간의 격차를 줄이기 위해서는 어휘 묶음뿐만 아니라 다른 다양한 언어적 특징들을 연구하여 좀 더 나은 학문적 글쓰기를 할 수 있도록 해야겠다고 말한다. 또한 덧붙여 지금까지 오랜 기간 동안 학습자가 일관성 있는 작문을 하기 위해 교수자가 하나의 단어나 두 개의 연어 정도만 가르치는 경우가 많았는데 많은 학문적인 글에서 다양한 어휘 묶음의 사용이 나타나는 만큼 그에 부합하여 학습자에게 이를 학습시키는 것이 중요하다고 이야기한다.

Pat Byrd와 Averil Coxhead(2010)는 'Lexical bundles in academic writing and in the teaching of EAP'에서 최근에 많은 연구가 이루어지고 있는 어휘 묶음에 대하여 여러 학문 영역에 걸쳐서 어떻게 나타나고 사용되고 있는지에 대해 연구하였다. 우선적으로 연구를 위해서 the Academic Word List(AWL) 코퍼스를 사용하였는데 이 코퍼스는 예술, 상업, 1법, 과학의 네 가지 학문 영역의 academic written English로 구성되어 있다. 또한 크게 네 가지의 학문 영역에 각각 7개씩의 하위 분야들로 나누어져 있는 것이 특징이다. 연구는 Wordsmith Tools 5.0을 사용하여 4단어로 이루어진 어휘 묶음을 확인하였다. 연구 결과 예술과 과학 분야에서 가장 적은 어휘 묶음이 확인되었고 반면에 법 분야에서 가장 많은 나타났고 상업 분야가 중간 정도로 나타났다. 또한 4가지 학문 분야에 걸쳐서 서로 공유하고 있는 어휘 묶음이 73개로 확인되었는데 이는 총 3백6십만 단어들 중에 39,616개의 단어들을 이루며 9,904번이 나타난 것으로 확인되었다. 즉 많은 빈도수로 나타나고 있는 이러한 어휘 묶음이 총 코퍼스 단어에서 1.1%의 비율을 차지하고 있었다. 많은 빈도수의 어휘 묶음을 보면 'on the basis of'가 380번으로 가장 많았고 'on the other hand'와 'as a result of', 'the end of the'등이 뒤를 잇는 것으로 나타났다. 또한 이 연구에서는 Biber, Conrad&Cortes(2004)와 Hyland(2008)와 같은 다른 학문적인 글에서 나타난 어휘 묶음과도 비교해보았다. 또한 기능과 구조에 따라서도 분류하여 확인해보았다. 이런 연구 결과들을 바탕으로 EAP(English for Academic Purposes) teaching에 있어서 교수자는 학생들에게 가르치는데 있어서 유용하게 사용할 수 있을 것이라고 주장한다.

또 Danica Joy Lorenzo Salazar(2011)는 'Lexical bundles in scientific English: A corpus-based study of native and non-native writing'에서 출판된 과학 분야 글들을 코퍼스로 이용하여 어휘 묶음을 빈도, 기능, 구조

에 따라 연구하였다. 이 연구를 위하여 생물학과 생화학, 생의학, 의학 분야에서 나온 원어민 작가의 논문들로 이루어진 Health Science Corpus(HSC)에서 총 백 삼 십 만개의 단어들 중 3개에서 6개의 단어로 이루어진 어휘 묶음을 추출하였다. 이 연구의 주된 목적은 HSC에서 가장 많이 나타나고 있는 어휘 묶음이 무엇인지 확인하고 이것들의 기능적, 구조적 특징들을 밝혀내어 이에 따라 분류해보고 이러한 어휘 묶음들이 비원어민의 글쓰기에서 나타나는지, 원어민과 비원어민간의 차이점은 무엇인지 살펴보는 것이다. 비원어민과의 비교를 위하여 사용될 코퍼스는 아르헨티나의 Rio Cuarto 국립 대학의 Iliana Martinez 교수가 수집한 생물학 분야의 스페인어 모국어자의 원고를 사용하였다. 연구 결과 빈도수를 보면 769개의 어휘 묶음이 나타났고 이는 전체에서 2% 정도를 차지한다. 이 중에서 3단어로 이루어진 어휘 묶음이 가장 많은 83%의 비율로 나타났고 4단어로 이루어진 것은 15%로 나타났다. ‘the presence of’와 ‘data not shown’ 이 가장 많은 빈도수로 나타났다. 구조상으로는 명사구와 of-phrase(전치사구)가 함께 쓰인 경우가 압도적으로 많은 빈도를 나타내었고 뒤를 이어 수동태와 전치사구가 함께 쓰인 경우가 많이 나타났다. 기능적으로는 text-oriented(텍스트 중심)의 어휘 묶음이 많이 나타났는데 그 중에서도 ‘were able to’, ‘it appears that’, ‘this suggests that’과 같은 inferential(추론) 기능의 어휘 묶음이 많이 나타났다. 다음으로 non-native expert corpus(영어비모국어 전문가 코퍼스, NNS)와 비교한 결과 442개의 target bundles(연구 대상 어휘 묶음) 가운데 312가 NNS에서 나타난 것으로 확인되었다. NNS에서 나타난 어휘 묶음의 빈도수로 ‘in the presence of’가 가장 많았고 뒤이어 ‘in order to’, ‘the number of’, ‘as well as’등이 나타났다. 또한 원어민에 비해 비원어민의 글에서 ‘were able to’, ‘were carried out’, ‘the fact that’, ‘in order to’, ‘in this study’등이 더 많이 사용된 것으로 확인되었고

반면에 ‘the ability of’, ‘as described previously’, ‘the possibility that’, ‘data not shown’ 등은 덜 사용된 것으로 나타났다. 구조나 기능적 면에서는 비슷하게 나타났다. 이러한 연구 결과들을 바탕으로 확인한 결과 이전에 다른 연구들과는 달리 어휘 묶음의 사용에 있어서 큰 차이점은 발견되지 않았고 오히려 유사한 점들이 훨씬 많이 나타났다. 원어민이 아니더라도 영어 교육을 많이 받은 사람들이 원어민에 가까운 실력을 보이고 있다고 볼 수 있겠다. 그럼에도 불구하고 이 연구 결과에서 나타난 비원어민이 원어민에 비해 어휘 묶음의 사용에 있어서 덜 다양하고, 자신 있고 많이 쓰는 표현들을 과도하게 사용하고 있는 점이나 기능면에서 participant-oriented(참여자 중심)적인 표현이 덜 쓰이는 점 등의 차이점의 발견은 고무적이라고 할 수 있겠다.

2.4 코퍼스 관련 수업 활용 연구

성일호(2009)는 ‘코퍼스 분석을 통한 조동사활용 수업 안 개발’에서 영어 표현에 있어 작가나 화자의 태도나 분위기를 잘 나타내주는 서법 조동사의 사용을 연구하여 이것을 가르치는데 있어서 효과적인 수업안을 개발하는데 목적을 두고 있다. 이 연구를 위해 The British National Corpus, The Cambridge International Corpus, Harry Potter series(해리포터 시리즈), Korean high school English textbooks(우리나라 고등학교 영어 교과서), Pop music lyrics(대중가요 가사), President Bush’s inauguration speech(부시 대통령 취임 연설문), Movie subtitles(영화 대본)을 코퍼스로 활용하였다. 특히 이 연구는 영어 교과서와 원어민들의 코퍼스를 비교하여 어떻게 서법 조동사가 쓰이고 있는지에 대해 관심을 가졌다. 프로그램으로는 Antconc를 이용하였다. 코퍼스 활용 조동사 수업의 형태는 과제 중심

교수법(Task-based approach) 형태로 영어 교육 전공 학생들 30명을 대상으로 한 학기동안 진행하였다. 우선 코퍼스 간 조동사 비율 분포를 보면 조동사 중 'will, would, can, could'의 사용이 가장 많음을 알 수 있었다. 또한 'shall, ought to'는 거의 사용되지 않았다. 또한 발견된 큰 특징으로는 우리나라 영어 교과서의 조동사 형태와 영어권 코퍼스 간에 큰 차이가 있는데 단순한 조동사별 빈도수에서 나아가 용법의 차이를 확인할 수 있었다. 'must'의 경우 영어교과서에서는 '의무'를 나타내는데 많이 쓰인 반면 영어권 코퍼스에서는 '예측'의 의미가 많이 나타났다. 그 외에도 여러 차이들이 나타났는데 이런 점들을 고려하여 영어 교과서 제작에 있어서 반영할 필요가 있을 것이다. 또한 이 연구에서는 이렇게 진행한 수업의 효과를 확인하기 위해 '조동사 이해도 테스트'를 만들어 사전, 사후 테스트를 실행하였다. 그 결과 어느 정도의 유의미한 결과 차이를 보임으로써 효과가 있음을 입증할 수 있었다. 이러한 결과들을 바탕으로 문법 이론을 실제 언어 사용과 결합시켜 학생들이 언어 사용 능력을 갖출 수 있도록 보다 효율적인 방안들을 제시할 수 있어야 한다고 주장한다.

또한 성일호(2010)는 '언어의 단어 결합력 측정 코퍼스 수업 안 개발'에서 원어민들이 어휘 덩어리로 의사소통을 하기 때문에 이를 이해하고 학습하는 것이 중요하므로 이러한 언어를 학습시키기 위하여 코퍼스를 이용한 효과적인 수업 안을 개발하는데 목적이 있다. 이 연구를 위하여 영어교육전공 4학년 40명을 대상으로 한 학기 동안 수업을 진행하였다. 통계를 활용한 어휘 결합력 측정에 대해 소개하며 통계 기법의 차이점도 함께 제시한다. 이렇게 어휘 결합력 측정 코퍼스 수업을 받은 학생들에게 수업을 받은 소감을 받은 후 다시 이를 정리하여 재설문지를 만든 후 학습자의 반응을 살핀 후 결과를 정리하였다. 조사 결과 실제 언어가 주는 코퍼스 자료는 동기유발과 읽기 학습의 상승효과를 가져오는 반면에 아직은 친숙치 않은 코퍼스 수업의 컴

퓨터 프로그램 운용과 통계 지식의 습득은 어려워하는 것으로 나타났다. 어휘, 문법, 쓰기 학습은 중간 위치를 차지하고 있지만 코퍼스 수업을 적절히 운용함에 따라 충분히 개선될 수 있는 가능성을 보였다. 이렇듯 코퍼스 수업이 연어 능력 시험에서 유의미한 차이가 있음이 확인됨으로써 학습자들이 지금까지 사전적 지식으로 단순히 숙어를 암기, 숙지하는 수준에서 단어의 연관성을 이해하는 것에서 벗어나 단어 간의 결합력 파악을 통계적 지식을 활용해 보다 객관적으로 측정하는 동안 연어에 대한 이해력이 높아졌음을 알 수 있었다. 어휘 덩어리를 구성하는 요소들을 어휘 문법의 개념으로 확장하고, 통계학적으로 검증해 사전이나 문법서를 편찬하는 것이 최근 추세인 만큼 이 연구를 통해 통계적 지식을 활용해 단어들의 결속력을 파악하는 수업 모형의 개발을 어휘 학습의 새로운 방안으로 제시하고자 하였다.

Ⅲ. 연구 방법 및 실제

3.1 연구 대상 및 자료수집

본 연구에서 원어민 화자와 한국인 화자의 영어 관련 분야 학술 논문 비교를 위해서 사용한 자료는 영어 관련 학술지에 게재된 원어민 화자와 한국인 화자의 학술 논문이다. 영어 관련 학문에 어느 한쪽으로 치우치지 않기 위해 영문학, 영어학, 영어교육학의 세 부분으로 나누어 자료를 수집하였다. 이렇게 세 분야로 나누어 연구한 이유는 같은 영어 관련 분야라고 하더라도 학문마다 글을 쓰는 형식이나 스타일이 다르고 주제가 확연히 다르기 때문에 그런 점에서 차이를 확인하고자 하였다. 코퍼스 구축을 위해 인터넷 학술 데이터 검색 프로그램을 이용하여 원어민 화자의 학술 논문 총 130개, 한국인 화자의 학술 논문 총 150개를 각각 수집하였다. 발표된 논문의 시기는 각각 너무 차이가 나면 비교에 있어서 문제가 있을 수 있으므로 최근 2008년에서 2012년 사이로 하였다. 원어민 화자의 경우 영어권 나라에서 발행되는 학술지에서 학자의 이름과 출신 학교 등을 확인하여 수집하였다. 물론 한국인 화자라고 하여도 원어민에 가까운 영어를 구사하는 경우도 있을 수 있고 학술 논문에 있어 원어민에 의한 개입이 있을 수 있는 문제가 있지만 모든 화자와 논문의 배경을 모두 파악하는 것은 다소 무리가 따르므로 최대한 데이터의 양을 충분히 하여 이러한 점을 해소하려 하였다. 또한 학술 논문을 수집한 학술지 선택의 경우 그 종류가 많지만 해당 분야의 대표적인 학술지 중 임의로 선택하였다. 해당 학술지의 출처와 각 학술지당 수집된 논문의 수는 다음 <표 1> 과 같다.

<표 1> 학술 논문 학술지 출처와 논문 수

원어민 화자 학술 논문 출처 (논문 수)	한국인 화자 학술 논문 출처 (논문 수)
Oxford-English (20개) Cambridge-English Language and Linguistics (20개) Oxford-English Language Teaching Journal (30개) Journal of English Linguistic (20개) American Literature (20개) English Education (20개)	새한영어영문학 (30개) 신영어영문학 (30개) 영어영문학 (27개) 현대영어영문학 (27개) 영어학 (36개)

영어 분야인 영문학, 영어학, 영어교육학의 어휘 수를 어느 한쪽으로 치우치지 않기 위해 최대한 분야 당 개수를 맞추어서 자료 수집을 하였다. 각 분야 당 논문 개수는 다음의 <표 2> 와 같다.

<표 2> 학술 논문의 영어 분야 당 논문 개수

분야	원어민 화자의 학술 논문 개수	한국인 화자의 학술 논문 개수
영문학	39개	49개
영어학	40개	50개
영어 교육학	51개	51개

이렇게 수집한 논문은 모두 PDF파일로 되어 있기 때문에 코퍼스 프로그램으로 분석하기 위해서 모두 텍스트 파일로 전환하였다. 전환 시 보다 정확한 데이터 분석을 위하여 학술 논문의 각주와 참조, 표와 같은 부분은 포함시키지 않고 논문 초록과 본문만을 텍스트 파일로 만들었다. 구축된 코퍼스는 원어민 화자의 학술 논문 코퍼스의 경우 총 어휘의 유형(type)은 36928

개, 총 어휘의 구현(token)은 943476개로 나타났고 한국인 화자의 학술논문 코퍼스의 경우 총 어휘의 유형은 26479개, 총 어휘의 구현은 779749개로 나타났다. 총 어휘 수에서 두 코퍼스 간에 163727개의 어휘 수가 차이가 나는 것은 처음에 자료를 수집할 때 각각 100개씩의 논문을 영어 학문 분야를 구분 짓지 않고 무작위로 수집한 후 어휘 수를 비교해 보니 한국인 화자의 학술논문의 어휘 수가 원어민 화자의 학술논문의 어휘 수에 비해 부족하여 어휘 수와 분야별 균형 또한 맞추기 위해 50개의 논문을 추가로 수집하였다. 그 후 다시 학문 분야별로 비교해 보니 원어민 화자의 학술논문에서 영어 교육 분야 논문이 부족하여 30개를 추가로 수집하여 균형을 맞추다보니 총 어휘 수에서는 차이가 나게 되었다.

3.2 연구 도구

본 연구를 하기 위해 주로 사용된 코퍼스 프로그램은 AntConc3.2.4w 이다. 이 프로그램은 초보자도 어렵지 않게 사용이 용이한 프로그램으로써 원래는 학생들이 수업하면서 사용할 수 있도록 개발되었으나 연구자들에게 많은 양의 텍스트를 분석할 수 있도록 해주는 도구로써 이용되고 있다. 일본 와세다 대학 교수 Laurence Anthony가 개발한 프리웨어 콘코던스 프로그램으로 Windows, Mac OS X, 리눅스 등과 같은 모든 운영체제에서 사용 가능하다. 주요 기능으로는 텍스트의 사용된 어휘의 목록과 빈도를 확인할 수 있는 Word List와 참조 텍스트와의 비교를 통해서 주요하게 사용된 어휘를 확인할 수 있게 해주는 Keyword List가 있다. 또한 특정 단어가 어떤 단어와 결합하여 사용되는지 확인해주는 Collocates나 연쇄적으로 하나의 덩어리로서 사용되는 표현들을 알아볼 수 있는 clusters, 어느 텍스트에서 얼마나 나타나는지 텍스트별로 빈도수나 분포를 확인할 수 있는 Concordance Plot,

확인된 어휘나 구 등을 텍스트에서 찾아 확인할 수 있는 File View 등의 기능들이 있다. 본 연구에서는 어휘 묶음 확인하고 비교하기 위하여 Clusters의 하위 기능인 N-Gram을 이용하였고 확인된 어휘 묶음들이 어떤 텍스트에서 얼마만큼의 빈도로 나타나는지 확인하기 위하여 Concordance Plot을 함께 사용하였다. 어휘들의 차이를 보기 위하여 Word List의 기능 역시 사용하였고 Keyword List나 Word List의 경우 AntConc3.2.4w 에서는 빈도만 확인이 되고 비율이 확인되지 않기 때문에 좀 더 정확한 데이터를 위하여 역시 대표적인 코퍼스 프로그램 중 하나인 WordSmith Tools 5.0의 Word list와 Keyword List를 함께 사용하였다.

3.3 연구 절차와 방법

본 논문의 연구 문제들을 해결하기 위하여 다음과 같은 절차와 방법을 사용하였다.

3.3.1 어휘 비교

우선 어휘 비교의 측면에서 보면 수집된 원어민 화자와 한국인 화자의 영어 분야 학술 논문을 텍스트 파일로 전환하여 위에서 말한 AntConc3.2.4w 와 WordSmith Tools 5.0을 이용하여 Word List를 뽑는다. 이 때 AntConc3.2.4w로 데이터를 확인할 시에는 모든 단어는 소문자로 인식하여 처리한다. 각각의 Word List에 나타난 어휘의 빈도에 따른 순위를 확인하고 전체 텍스트에서 나타난 비율을 각각 어떠한지 확인한다. 이를 통해 본 연구를 통해 확인하고자 하는 원어민 화자와 한국인 화자의 학술 논문에 나타난 기능어들의 사용, 대명사, 서법 조동사 접속(부)사와 같은 어휘들의 사용 빈

도와 비율을 알아볼 수 있다. 특히 접속(부)사의 문두, 문중 위치 비교에 있어서는 문두에서는 대문자로 나타나고 문중에서는 소문자로 나타나기 때문에 모든 단어를 소문자로 인식하여 처리하게 하지 않고 대소문자를 구분하여 빈도를 확인할 수 있도록 하여 확인한다.

3.3.2 어휘 묶음 비교

다음으로 어휘 묶음을 확인하기 위해서는 AntConc3.2.4w 프로그램의 clusters의 메뉴를 이용한다. 연속된 어휘들의 사용 빈도를 알아보기 위해 하위 메뉴인 N-Gram을 이용하여 cluster의 크기를 4개로 지정하여 가장 많이 나타나고 유용하게 사용할 수 있는 4단어로 이루어진 어휘 묶음을 확인한다. 많은 어휘 묶음이 있겠지만 많이 사용하는 것이 유의미 하므로 나타나는 빈도수는 Min. N-Gram Frequency를 15로 지정하여 전체 텍스트에서 최소 15번 이상 나타나는 것으로 한다. 나타난 어휘 묶음은 각각의 분야별로 나누어 몇 개의 텍스트에서 얼마의 빈도수로 나타나는지 확인한다. 다음으로 나타난 어휘 묶음을 기능과 구조적인 측면으로 나누어 각각의 차이를 살펴본다.

3.3.3 중요 어휘 비교

두 코퍼스 간에 중요 어휘를 분석하기 위해서 우선적으로 참조 코퍼스가 필요한데 BNC(British National Corpus) 코퍼스를 사용한다. 프로그램은 WordSmith Tools 5.0을 사용한다. 가장 보편적으로 많이 사용되는 어휘를 확인할 수 있는 BNC 코퍼스를 활용하여 원어민 화자와 한국인 화자의 학술 논문에 나타난 두드러지고 중요하게 나타나는 어휘를 비교해 본다. 또한 원

어민 화자와 한국인 화자의 학술 논문 코퍼스 각각을 서로의 참조 코퍼스로 활용하여 Keyword List를 확인함으로써 서로의 코퍼스와 비교했을 때 두드러지는 어휘의 차이가 있는지 또한 확인해 본다.

IV. 연구 결과 및 논의

4.1 어휘 목록과 기능어 사용의 비교 분석

4.1.1 어휘 목록 비교 분석

언어 학습에 있어서 가장 기초가 되는 것은 어휘이다. 언어를 읽고, 듣고, 쓰고, 말하는데 있어서 어휘는 가장 기본적으로 필요한 요소이기 때문이다. 따라서 오늘날 영어 교육에 있어서도 어휘 학습이 강조되고 있는 것이 사실이다. 이러한 어휘는 크게 명사, 동사, 형용사, 부사와 같이 단어 자체에 독자적인 의미를 지니고 있는 내용어(content word)와 관사, 대명사, 전치사, 접속사, 조동사와 같이 실질적으로 그 단어의 의미를 나타내기보다 단어와 단어를 연결해 주는 문법적 기능만을 하는 기능어(function word)로 나눌 수 있다. 내용어의 경우 의미를 가지기 때문에 우리는 내용어만 보더라도 많은 것을 파악할 수 있게 된다. 따라서 본 논문에서 가장 먼저 수집된 두 코퍼스 간의 어휘 목록을 비교해 보겠다. 어휘 목록에서 상위를 차지하는 기능어는 뒤에서 따로 다룰 것이기 때문에 내용어 중심으로 살펴보겠다. 해당 비교표는 다음의 <표 3> 과 같다. 참고로 앞으로 본 논문에서 English-Native는 EN으로, English-Non-native는 KN으로 줄여서 사용하겠다.

<표 3> 두 코퍼스 간의 어휘 목록 비교

N	English-Native			English-Non-native		
	어휘	빈도	비율	어휘	빈도	비율
1	Students	2464	0.25	English	3169	0.41
2	English	2372	0.24	Students	1949	0.25
3	Teachers	1779	0.18	Language	1640	0.21
4	Language	1666	0.17	Learners	1452	0.19
5	Social	1408	0.15	Korean	1424	0.19
6	Use	1364	0.14	Study	1169	0.15
7	Time	1161	0.12	Words	1101	0.14
8	Writing	1055	0.11	Different	1052	0.14
9	Example	1054	0.11	Learning	1012	0.13
10	New	1022	0.11	Writing	973	0.13
11	Study	1012	0.11	Subject	863	0.11
12	Data	981	0.1	Use	800	0.1
13	See	943	0.1	Reading	797	0.1
14	Work	936	0.1	Verb	766	0.1
15	Class	917	0.1	New	764	0.1
16	Reading	902	0.09	Group	763	0.1
17	Different	896	0.09	World	738	0.1
18	Teaching	887	0.09	Analysis	733	0.1
19	Words	818	0.08	Time	721	0.1
20	People	790	0.08	Class	692	0.09
21	School	765	0.08	Feedback	655	0.09
22	Learning	730	0.08	EFL	631	0.08
23	Subject	720	0.07	American	611	0.08
24	Research	701	0.07	Number	611	0.08
25	Form	697	0.07	Structure	606	0.08
26	Meaning	693	0.07	Research	605	0.08
27	Speakers	692	0.07	Sentence	587	0.08
28	Linguistic	690	0.07	Human	585	0.08
29	Analysis	679	0.07	Man	581	0.08
30	Classroom	654	0.07	Life	579	0.07
31	World	629	0.07	Teacher	579	0.07
32	Case	626	0.06	Level	567	0.07

33	Particular	618	0.06	Speakers	566	0.07
34	Poetry	611	0.06	Word	557	0.07
35	Learners	596	0.06	Native	553	0.07
36	American	581	0.06	Features	547	0.07
37	Speech	581	0.06	Meaning	547	0.07
38	Change	565	0.06	Verbs	534	0.07
39	Context	559	0.06	Process	529	0.07
40	Education	555	0.06	Test	523	0.07

전반적으로 빈도수 500번 이상으로 전체 어휘 목록에서 상위에 나타난 내용어 중심의 어휘들을 살펴보면 아무래도 같은 분야의 내용을 다루다보니 비슷하게 나오는 어휘들이 많이 있다. ‘English, student, language, learning, learners, words, writing, speakers’ 와 같은 어휘들은 빈도수에서의 약간의 차이는 있지만 상위 목록에서 함께 확인되었다. 동사로는 EN에서는 ‘use, see, change’등이 많은 빈도수로 나타났고 KN에서는 상위에서는 ‘use’만 나타났다. 형용사로는 EN에서는 ‘social, new, different, particular’등이 나타났고 KN에서는 ‘new’와 ‘different’만 함께 나타났다. 글의 종류가 학술 논문이다 보니 두 코퍼스 모두 ‘example, data, subject, study, research, analysis, case’와 같은 어휘들이 상위에 많이 나타났다. 그리고 원어민 화자와 한국인 화자의 학술 논문이다 보니 EN에서는 ‘American’이, KN에서는 ‘Korean’과 ‘native’의 어휘가 각각 상위에 확인되었다. 그 외에도 ‘school, linguistic, poetry, speech, context, education’ 등은 EN에서만 상위에서 나타났고, ‘group, feedback, EFL, human, man, life, level, test’등은 KN의 상위에서만 나타났다. 이렇듯 두 코퍼스 간의 많은 빈도수로 상위에 나타난 내용어들을 통하여 전체적으로 많이 쓰이는 명사, 동사, 형용사 등의 차이를 알아볼 수 있었다.

4.1.2 기능어 사용의 비교 분석

시간이 지날수록 어휘가 계속적으로 생겨나 그 양이 방대해지는 내용어와는 달리 기능어는 기존의 어휘에서 더 이상 확장되지 않는 적은 수로도 언어 사용에 있어 중요한 역할을 하고 있다. 따라서 기능어의 사용이 제대로 이루어지지 않는다면 단순 어휘의 나열일 뿐 매끄러운 언어 사용이 이루어질 수 없다. 그만큼 영어의 사용에 있어서 기능어의 적절한 사용은 무엇보다 강조되어야 할 것이다. 그렇다면보니 코퍼스를 이용해 어휘 목록을 뽑아 보면 다양한 기능어들이 많은 빈도수로 상위 목록에서 확인되어진다. 그러한 측면에서 이 논문에서도 수집한 두 코퍼스 간의 어휘 목록에서 각각 기능어(function word)의 사용 빈도만 따로 비교해 보았다. 각각의 어휘 목록에서 상위에 나타난 20개의 기능어와 그 사용 빈도수와 비율을 나타낸 표는 다음의 <표 4> 와 같다.

<표 4> 기능어의 사용 빈도수 & 비율

N	English-Native			English-Non-native		
	어휘	빈도	비율	어휘	빈도	비율
1	the	57799	6	the	53268	6.9
2	of	37434	3.9	of	29739	3.9
3	and	27732	2.9	and	21497	2.8
4	in	25088	2.6	in	20935	2.7
5	to	23676	2.5	to	17677	2.3
6	a	19346	2	a	14385	1.9
7	that	13454	1.4	is	11299	1.5
8	is	12266	1.3	that	9461	1.2
9	as	10721	1.1	as	8150	1.1
10	for	8271	0.9	for	5658	0.7
11	with	6865	0.7	with	5107	0.7

12	this	6324	0.7	this	4303	0.6
13	it	5857	0.6	by	4251	0.6
14	on	5246	0.5	are	4233	0.6
15	are	5069	0.5	on	4213	0.5
16	be	4897	0.5	be	4199	0.5
17	by	4814	0.5	not	3992	0.5
18	not	4615	0.5	it	3873	0.5
19	or	4380	0.5	his	3699	0.5
20	from	4132	0.4	or	3128	0.4

결과를 보면 두 코퍼스 간의 기능어 빈도가 전체적으로 비슷하게 나타나는 것을 확인할 수 있다. 비율은 6.9%로 EN이 조금 높지만 정관사 ‘the’가 가장 많이 사용되고 있다. 5만번 이상 나타나는 것으로 압도적인 빈도수를 나타낸다. 다음으로 전치사 ‘of’는 3.9%의 같은 비율로 확인되었다. 뒤이어 접속사 ‘and’, 전치사 ‘in, to’ 부정관사 ‘a’가 비슷한 비율로 나타났다. 그 뒤로도 순위만 약간씩 차이가 날 뿐 비슷하게 접속사 ‘as’, 전치사 ‘for, with’, 대명사 ‘this’등이 나타났다. 차이점이라고 한다면 EN에서는 전치사 ‘from’이 좀 더 많은 빈도수로 나타났고 KN의 경우 대명사 ‘his’가 좀 더 많은 빈도수로 나타났다. 전반적으로 1억 단어 이상의 단어로 구축되어 가장 널리 사용되는 코퍼스 BNC와 비교해 봤을 때 ‘the, and, to, that, it, a, of’ 등과 같은 기능어가 상위 목록에서 확인되는 것은 비슷하게 나타나는 반면 아무래도 본 논문의 코퍼스가 논문으로 이루어져 있기 때문에 BNC의 상위에서 나타나는 대명사 ‘you, we, they’ 등은 나타나지 않았다.

이렇듯 기능어는 각종 코퍼스에서 상위에 많은 빈도수로 나타나는 만큼 영어 교육에 있어서 내용이 못지않은 중요성을 지님을 인식시키고 학습시킴으로써 자연스럽게 매끄러운 영어를 사용할 수 있도록 하는 것이 무엇보다 필요할 것이다.

4.2 대명사와 서법조동사 사용의 비교 분석

4.2.1 대명사 사용의 비교 분석

자신의 연구에 대한 내용을 정확하고 조리 있게 전달하기 위한 목적을 가진 학술 논문을 작성하는데 있어서 다양한 기능어 중에서도 무언가를 지칭해 주는 역할을 하는 다양한 대명사의 사용은 중요하고 따라서 많이 나타나게 된다. 가리키는 것이 사람이든 사물이든 글쓴이가 정확한 대명사를 사용하지 않으면 글을 읽는 사람들은 내용을 파악하기가 힘들어지게 된다. 따라서 본 논문에서 두 코퍼스 간에 대표적으로 많이 쓰이는 인칭대명사와 지시대명사를 중심으로 그 사용 빈도와 비율을 비교해 보았다. 그 결과는 다음의 <표 5> 와 같다.

<표 5> 대명사의 사용 빈도수 & 비율

대명사	English-Native		English-Non-native	
	빈도	비율	빈도	비율
I	3511	0.36	2142	0.28
We	3330	0.35	1241	0.16
You	1033	0.11	524	0.07
They	3048	0.32	2216	0.29
He	2464	0.26	2245	0.29
She	1766	0.18	1042	0.14
It	5857	0.61	3873	0.5
This	6324	0.66	4303	0.56
That	13454	1.4	9461	1.23
These	2530	0.26	968	0.13
Those	804	0.08	511	0.07
Him	554	0.06	545	0.07

Her	2671	0.28	1835	0.24
Them	1142	0.12	778	0.1
Us	702	0.07	394	0.05
Me	482	0.05	309	0.04

결과를 보면 우선적으로 사람을 지칭하는 인칭대명사를 비교해 보면 ‘I, they, he’의 경우는 비율적으로 큰 차이를 보이지 않고 비슷하게 나타나는 것을 확인할 수 있다. 반면에 차이를 보이는 것은 ‘we’와 ‘you’인데 먼저 ‘we’의 경우 EN의 코퍼스에서는 빈도수가 3330번에 비율이 0.35%인데 KN의 경우 빈도수가 1241번에 0.16%로 나타났다. 거의 두 배 이상의 차이가 나타나는 것을 알 수 있는데 이는 논문을 작성한 저자가 공동연구로 인한 ‘we’의 사용이 더 많은 것인지 생각해 볼 수 있기 때문에 각각 수집한 논문의 공동연구가 얼마나 되는지 확인을 해 보았다. EN의 논문의 경우는 총 130개의 논문 중에 공동연구 논문이 36개로 확인되었고 KN의 경우는 총 150개의 논문 중에서 26개로 확인되었다. 이것으로 보아 ‘we’의 사용이 어느 정도 차이는 날 수 있겠지만 두 배 이상의 차이가 날 정도의 논문 수 차이는 아니기 때문에 글을 쓰는 데 있어서의 사고방식이나 스타일의 차이가 있는 것으로 생각되어진다. EN의 논문에서 ‘I’와 ‘we’의 사용 빈도가 차이가 거의 없는 반면 KN에서는 거의 두 배 가까이 차이가 나는 것은 분명한 차이임은 확실하다. 하지만 이러한 차이가 영어영문학 학술 논문이 아닌 다른 유형의 글에서도 나타날 것이라고 일반화하는 것은 다소 무리가 있을 수 있다. 그러한 결론의 일반화를 위해서는 좀 더 많은 연구가 필요하다고 판단된다.

다음으로 사물을 지칭하는 지시대명사를 보면 두 코퍼스 모두 ‘it’과 ‘this’가 비슷한 비율로 많이 나타나고 있다. ‘that’의 경우도 많이 나타나고 있지만 이는 ‘that’이 지시대명사가 아닌 관계대명사로도 많이 사용되고 있기 때

문에 다른 어휘들보다 훨씬 많이 나타나는 것으로 보여 진다. ‘these’와 ‘those’의 경우 EN의 논문에서는 ‘these’가 ‘those’보다 세 배 이상 사용된 것으로 나타난 반면, KN 논문에서는 단지 두 배 정도 더 많이 사용된 것으로 나타났다. 더 나아가 ‘these’의 경우 KN의 논문에 비해 EN의 논문에서 두 배 이상 더 많이 사용된 것으로 나타났다. 그리고 ‘him, her, them, us, me’ 와 같은 목적어로 많이 사용되는 대명사의 경우 두 코퍼스’의 간에 거의 비슷한 비율로 나타난 것으로 확인되어 뚜렷한 차이는 보이지 않았다. ‘her’의 경우 다른 목적어에 비해 사용 빈도가 높는데 이는 ‘her’가 목적격과 소유격 모두에서 쓰이기 때문에 더 많이 나타날 수밖에 없고 본 논문에서는 나타난 어휘에 초점을 맞추므로써 어떠한 기능으로 사용되었는지의 구분은 따로 하지 않았다.

4.2.2 서법조동사 사용의 비교 분석

영어의 사용에 있어서 서법조동사의 사용은 서로 간에 정확한 의사소통을 위해서 중요한 역할을 한다. 어떠한 조동사를 사용하느냐에 따라서 전달하고자 하는 내용이나 뉘앙스가 상당히 달라질 수 있기 때문이다. 그렇기 때문에 EFL 환경에서 영어를 배우는 우리나라 영어 학습자들에게 있어서 서법조동사의 다양한 의미와 쓰임, 뉘앙스 차이 등을 익혀 원어민처럼 구사하는 것이 결코 쉬운 일은 아닌 것이다. 따라서 서법조동사의 사용에 있어 원어민과 우리나라 영어 학습자간의 비교 연구는 중요하다고 할 수 있다. 그리고 서법조동사는 어떠한 텍스트의 종류와 상관없이 보편적으로 사용되고 그 용법이 다양하기 때문에 언어의 사용 양상이나 특성을 알기에 용이하다. 따라서 이 논문에서도 두 코퍼스 간의 서법조동사 사용 빈도를 살펴보고 다른 연구에서의 결과와도 비교해 보려한다. 우선적으로 본 논문에서의 결과는 다음의

<표 6> 과 같다.

<표 6> 서법조동사의 사용 빈도수 & 비율

서법 조동사	English-Native		English-Non-native	
	빈도	비율	빈도	비율
can	2174	0.23	1937	0.25
could	964	0.1	492	0.06
may	1262	0.13	825	0.11
might	806	0.08	330	0.04
will	1117	0.12	809	0.11
would	1431	0.15	573	0.07
should	515	0.05	691	0.09
must	467	0.05	335	0.04
shall	116	0.01	52	
cannot	308	0.03	366	0.05
합계	9160	1.02	6410	0.82

결과를 보면 전체 코퍼스에서 모든 서법조동사의 사용비율이 EN의 경우 1.02%로 나타났고 KN의 경우 0.82%로 나타났다. 전체적인 사용에 있어서 눈에 띄는 큰 차이를 보이는 것은 아니지만 원어민의 경우가 한국인 화자에 비해 좀 더 서법조동사의 사용이 많음을 알 수 있다. 또한 두 코퍼스 모두 'can'의 사용이 가장 많은 것으로 확인되었다. 서법 조동사 'can, may, will'과 같은 현재형이 'could, might, would'보다 많이 사용되고 있음을 알 수 있는데 이는 고광윤(2008)의 논문에서 영어교과서와 Frown 코퍼스에서 서법조동사 사용빈도 비교에서 나타난 결과와도 같음을 알 수 있다. 이 논문에서는 이러한 결과는 영어교과서에서 논리적 추론보다는 능력을 나타내는 예문이 많고 더불어 추론을 나타낼 때도 'can'을 더 선호하는 경향과 관계있을 것이라고 밝혔다. 이와 마찬가지로 본 논문에서도 두 코퍼스 모두

‘may’보다 ‘can’이 두 배 이상 빈도수가 나타나는 것으로 확인되었는데 그 결과와 일맥상통하는 부분이 있다고 생각된다.

좀 더 자세히 ‘can’과 ‘could’의 사용을 비교해 보면 EN의 경우 ‘can’이 0.23%, ‘could’가 0.1%로 사용 빈도에서 두 배 정도의 차이를 보이는 반면 KN의 경우는 ‘can’이 0.25%, ‘could’가 0.06%로 네 배 이상의 차이를 나타내는 것으로 확인되었다. 원어민에 비해 한국인 화자의 ‘could’사용이 더 적음을 알 수 있다. ‘may’와 ‘might’의 사용을 비교해 보면 EN의 경우 ‘may’는 각각 0.13%, 0.11%로 비슷하게 나타났지만 ‘can’과 ‘could’에서처럼 ‘might’에서 차이를 보였다. 각각 0.08%, 0.04%로 나타나 두 배의 차이를 보이고 있다. 이 역시도 원어민 화자에 비해 한국인 화자가 ‘might’의 사용이 적음을 알 수 있다. 이는 또한 ‘will’과 ‘would’에서도 마찬가지로 나타났다. ‘will’의 경우 각각 0.12%, 0.11%로 비슷하게 나타났으나 ‘would’의 경우 각각 0.15%, 0.07%로 두 배 이상의 차이가 나타나는 것으로 확인되었다. 이 역시도 원어민 화자에 비해 한국인 화자가 ‘would’의 사용이 적음을 알 수 있다. 이 결과만으로 단정 지을 수는 없겠지만 분명하게 서법 조동사의 과거형의 사용에서 큰 차이를 보임으로써 한국인 화자가 서법조동사의 사용에 있어서 현재형을 좀 더 선호하고 과거형의 사용에서 다소 어려움을 느낀다고 생각되어 진다.

의무와 확신을 나타내는 서법조동사 ‘should’와 ‘must’의 경우를 보면 EN의 경우 둘 다 0.05%의 비율로 비슷하게 사용된 반면 KN의 경우 ‘should’가 0.09%, ‘must’가 0.04%로 두 배 이상 차이가 남을 확인할 수 있다. 이로써 원어민 화자에 비해 한국인 화자가 좀 더 강한 의미의 ‘must’보다는 편하게 보편적으로 쓸 수 있는 ‘should’의 사용을 더 선호하는 것으로 생각되어 진다. 더불어 전반적으로 많이 나타나지 않는 ‘shall’의 경우에서도 EN에서 두 배 정도 더 나타난 것을 확인할 수 있었다.

이렇듯 서법조동사의 빈도수를 확인해 본 결과 전체적인 서법조동사의 사용에서도 EN의 경우가 많았고 ‘could, might, would’와 같은 서법 조동사의 과거형의 사용도 역시 많은 것을 확인한 만큼 우리나라 영어 교육에 있어서 어느 한쪽으로 치우쳐 사용하기보다 좀 더 다양한 조동사를 활용하여 의미전달을 할 수 있도록 교육하는 것이 반드시 필요할 것으로 생각된다.

4.3 어휘 묶음 사용의 비교 분석

어휘 묶음(lexical bundle)이란 확장된 연어(collocation)로써 특정 사용역(register)에서 통계적으로 자주 함께 나타나는 3개 혹은 그 이상의 단어들의 연쇄체라고 할 수 있다. 많은 어휘 묶음은 우리가 일반적으로 알고 있는 관용어(idiom)와 달리 어휘 묶음을 구성하는 단어들 자체만으로 그 의미를 파악할 수 있다. 이러한 고정된 단어들의 연쇄체의 사용은 언어를 사용하는데 있어서 상당히 중요한 역할을 한다. 얼마나 잘 알고 사용하느냐에 따라서 원어민과 같이 유창한 영어 구사를 할 수 있는 지름길이 될 수 있기 때문이다. 그렇기 때문에 개별적인 어휘들을 잘 익히고 사용하는 것도 중요하지만 여기서 한 단계 더 나아가 어휘 묶음을 익히고 학습하는 것이야말로 영어 사용에 있어서 무엇보다도 필요한 부분이라고 할 수 있다. 따라서 본 논문에서 학술 논문에 나타난 어휘 묶음의 종류와 사용 양상을 살펴보고 원어민 화자와 한국인 화자의 학술 논문사이에 나타난 차이점을 알아보고 앞으로의 학문적 글쓰기에서 어휘 묶음의 사용에 도움이 되고자 한다. 본 논문에서는 두 코퍼스에서 나타나고 있는 다양한 어휘 묶음을 전체적으로 학문별 빈도와 나타난 텍스트 수를 알아보고 많이 나타나고 있는 어휘 묶음을 살펴보았다. 또 각각의 학문별로 많이 나타나고 있는 어휘 묶음을 따로 살펴보았다. 마지막으로 기능과 구조에 따라서도 분류해 보았다. 우선 어휘 묶음

을 학문별로 나누어 빈도를 살펴볼 것이기 때문에 수집된 코퍼스의 학문별 어휘 수를 확인해 보았다. 그 결과는 다음의 <표 7> 과 같다.

<표 7> 학문별 어휘 유형(type) & 구현(token) 수

분야	English-Native		English-Non-native	
	유형	구현	유형	구현
영문학	23304	312751	19423	282605
영어학	18136	342575	10192	217379
영어교육학	14517	288150	9626	279765

4.3.1 어휘 묶음의 사용 빈도 비교 분석

가장 먼저 두 코퍼스에서 빈도수 15번 이상 나타난 어휘 묶음을 빈도수와 몇 개의 텍스트에서 나타났는지 확인하고 각각 영문학, 영어학, 영어교육의 세 분야로 나누어 어느 분야에서 더 많이 나타나고 있는지 살펴보았다. 어휘 묶음의 나열은 찾아보기 쉽도록 알파벳순으로 나열하였다. 그 결과는 다음의 <표 8>, <표 9> 와 같다.

<표 8> English-Native의 학문별 어휘 묶음 빈도 & 텍스트 수

N	어휘 묶음	빈도	텍스트	영문학		영어학		영어교육	
				빈도	텍스트	빈도	텍스트	빈도	텍스트
1	a great deal of	25	20	4	4	12	9	9	7
2	a result of the	16	13	4	4	11	8	1	1
3	a wide range of	16	15			10	9	6	6
4	are more likely to	17	13	1	1	10	8	6	4
5	as a means of	21	18	11	8	5	5	5	5
6	as a result of	45	30	15	9	17	13	13	8
7	as part of a	23	18	10	8	4	4	9	6
8	as part of the	24	22	7	6	8	7	9	9
9	as readers and writers	19	2					19	2
10	as shown in figure	15	11			14	10	1	1
11	as we have seen	22	17	9	7	10	7	3	3
12	as well as in	17	13	3	2	12	9	2	2
13	as well as the	53	35	21	13	18	11	14	11
14	at the beginning of	25	17	6	3	2	2	17	12
15	at the end of	39	29	13	11	5	5	21	13
16	at the heart of	20	18	11	9	4	4	5	5
17	at the level of	15	12	3	3	10	7	2	2
18	at the same time	94	54	36	19	29	18	29	17
19	at the time of	40	27	5	4	18	13	17	10
20	can be seen in	15	10	1	1	12	8	1	1
21	can be used to	20	11	1	1	11	7	8	3
22	despite the fact that	16	10	7	3	9	7		
23	i would like to	21	14	3	2	11	6	7	6
24	in a variety of	17	14	1	1	10	7	6	6
25	in a way that	29	20	15	9	2	2	12	9
26	in late middle english	18	6			18	6		
27	in light of the	15	10	5	1	9	8	1	1
28	in other words the	17	13	3	2	5	4	9	7

29	in relation to the	28	21	8	6	12	8	8	7
30	in response to the	17	12	4	4	6	4	7	4
31	in terms of the	31	18	3	3	24	12	3	3
32	in the case of	83	40	17	9	57	24	9	7
33	in the context of	57	36	19	10	23	15	15	11
34	in the development of	16	11	2	2	7	4	7	5
35	in the face of	22	14	13	8	3	3	6	3
36	in the first place	20	15	10	7	8	6	2	2
37	in the form of	31	23	13	8	8	6	10	9
38	in the next section	17	13	1	1	14	10	2	2
39	in the north and	18	2			17	2		
40	in the present study	15	10			11	6	4	4
41	in the process of	29	16	2	2	3	3	24	11
42	in the role of	16	1					16	1
43	in the same way	19	10	5	2	3	3	5	5
44	in the speech of	21	9			20	9		
45	in the united states	57	27	22	9	15	7	19	11
46	in the use of	21	14	2	2	13	7	5	5
47	is likely to be	16	11			13	8	3	3
48	is one of the	21	18	7	5	8	7	6	6
49	it is clear that	34	20	11	8	20	10	3	2
50	it is difficult to	23	18	4	3	13	9	6	6
51	it is important to	45	27	4	4	18	13	23	10
52	it is necessary to	16	12	6	4	7	5	3	3
53	it is not the	15	12	7	6	5	4	2	2
54	it is possible that	17	13			14	10	3	3
55	it is possible to	25	18	8	6	11	7	5	5
56	it should be noted	16	13			12	10	4	3
57	of the relationship between	15	12	4	4	4	4	7	4
58	of the twentieth century	23	11	9	5	12	4	2	2
59	of the united states	15	10	9	4	3	3	3	3
60	of the white elephant	21	1	21	1				

61	on the basis of	43	28	2	2	31	16	10	10
62	on the one hand	39	24	23	15	14	7	2	2
63	on the other hand	73	43	18	14	47	22	8	7
64	on the part of	30	19	6	6	16	8	8	5
65	one of the most	28	25	10	9	10	9	8	7
66	over the course of	25	20	9	8	2	2	14	10
67	per cent of the	16	11			6	6	10	5
68	teaching for social justice	20	4					20	4
69	that there is a	30	19	4	4	19	12	4	3
70	that there is no	22	16	4	3	12	10	3	3
71	the beginning of the	29	23	10	8	10	7	9	8
72	the case of the	19	15	6	5	10	8	2	2
73	the context of the	18	15	7	4	7	7	4	4
74	the end of the	65	41	30	21	12	7	23	13
75	the extent to which	34	23	8	6	6	4	20	13
76	the fact that the	33	23	10	8	19	11	4	4
77	the first half of	16	11	4	4	11	6	1	1
78	the history of the	16	6	2	1	6	4	8	1
79	the language of the	17	11	5	4	10	5	2	2
80	the majority of the	20	11	1	1	4	4	15	6
81	the nature of the	31	20	5	4	20	11	6	5
82	the new english education	17	2					16	2
83	the new york times	16	10	14	9	2	1		
84	the part of the	17	14	3	3	7	7	6	4
85	the rest of the	19	16	3	3	12	10	4	3
86	the role of the	17	11	4	3	4	3	8	5
87	the social meaning of	18	2			18	2		
88	the source of the	25	4	1	1	24	3		
89	the time of the	19	11			8	4	11	7
90	the use of the	34	20	8	4	16	9	10	7
91	the way in which	17	12	3	2	6	5	8	5
92	the ways in which	62	25	10	8	7	5	45	12

93	to account for the	17	13	1	1	13	11	1	1
94	to do with the	16	10	8	6	7	4		
95	to the extent that	18	13	6	5	12	8		
96	ways in which the	16	9	2	2			13	7
97	what it means to	19	12	7	5	1	1	11	6
		2505	1572	640	433	1079	666	758	473

<표 9> English-Non-native의 학문별 어휘 묶음 빈도 & 텍스트 수

N	어휘 묶음	빈도	텍스트	영문학		영어학		영어교육	
				빈도	텍스트	빈도	텍스트	빈도	텍스트
1	an important role in	15	12	6	5	1	1	8	6
2	and at the same	20	10	11	7	2	2	1	1
3	are more likely to	15	8			8	4	7	4
4	argument of the verb	18	3			17	3		
5	as a result of	31	19	13	9	6	5	12	5
6	as one of the	23	15	4	3	3	3	11	9
7	as shown in table	46	20			4	2	42	18
8	as shown in the	19	13	1	1	6	5	8	7
9	as well as the	49	35	26	16	8	7	15	12
10	at the beginning of	36	24	9	7	4	3	23	14
11	at the end of	63	37	28	18	6	5	29	14
12	at the same time	68	44	35	20	13	11	20	13
13	be one of the	16	15	8	7			8	8
14	between the two groups	25	15	2	2	1	1	22	12
15	can be said that	16	10	4	4	2	2	10	4
16	can be seen as	15	12	10	7	2	2	3	3
17	difference between the two	17	13	2	2	2	2	13	9
18	different from that of	15	14	5	4	5	5	5	5
19	does not seem to	17	15	4	4	9	7	4	4

20	efl context of korea	22	2					22	2
21	for the purpose of	19	15	4	4	4	3	11	8
22	in addition to the	22	20	8	7	11	10	3	3
23	in relation to the	17	14	9	7	2	2	6	5
24	in sons and lovers	17	2	17	2				
25	in spite of the	19	15	6	6	10	7	3	2
26	in terms of the	54	34	10	6	17	11	27	17
27	in the case of	57	32	7	6	35	16	15	10
28	in the context of	21	14	4	2	9	7	8	5
29	in the efl context	20	4					20	4
30	in the form of	37	26	24	14	8	7	5	5
31	in the present study	16	10			1	1	15	9
32	in the previous section	18	12			17	11	1	1
33	in the process of	54	30	18	12	23	7	13	11
34	in the same way	25	17	8	6	12	8	5	3
35	in the sense that	31	19	8	5	16	8	7	6
36	in the target language	18						18	8
37	in the united states	18	13	8	6			10	7
38	in the use of	17	8	1	1	6	3	10	4
39	is based on the	19	15	3	3	9	5	7	7
40	is one of the	44	31	9	7	14	11	21	13
41	it can be said	16	10	4	4	3	3	9	3
42	it is important to	17	15	7	6	3	3	7	6
43	it is necessary to	17	16	7	6	3	3	7	7
44	it is not the	16	12	11	8	5	4		
45	it is true that	18	11	11	6	5	3	2	2
46	korean learners of english	25	12			8	4	17	8
47	native speakers of english	43	19			12	6	31	13
48	of the nineteenth century	17	5	16	5				
49	of the present study	15	12			1	1	14	11
50	of the target language	17	10			1	1	16	9
51	of the united states	16	6	13	5	3	1		

52	of the working class	16	3	14	3				
53	of this paper is	25	21	8	6	15	13	2	2
54	of this study is	25	21	2	2	8	6	15	13
55	on the basis of	53	31	9	5	25	15	19	11
56	on the other hand	172	80	27	15	87	36	58	29
57	one of the most	50	35	15	12	9	6	26	17
58	that there is no	24	16	3	3	16	10	3	3
59	the basis of the	15	11	5	4	7	5	3	2
60	the beginning of the	45	28	20	12	2	1	23	15
61	the content of the	15	11	1	1	5	3	9	7
62	the efl context of	21	3					21	3
63	the end of the	64	35	29	16	3	3	32	16
64	the fact that the	52	33	18	10	21	12	13	11
65	the house of the	19	2	18	2				
66	the meaning of the	20	9	9	3	8	5	3	1
67	the presence of a	17	6	1	1	14	4	1	1
68	the purpose of the	15	12	1	1	1	1	13	10
69	the purpose of this	28	24	6	4	11	11	11	9
70	the relationship between the	17	14	8	7	2	2	7	5
71	the result of the	18	14	2	2	6	5	10	7
72	the results of the	51	25	2	2	6	6	43	17
73	the results showed that	16	10			2	2	14	8
74	the structure of the	23	11	14	6	4	1	4	4
75	the subject of the	20	10	2	2	15	5	3	3
76	the total number of	15	11			3	2	12	9
77	the turn of the	23	4	23	4				
78	the two types of	26	5			3	2	23	3
79	the use of the	20	9			7	3	13	6
80	the ways in which	21	15	14	8	2	2	5	5
81	this paper is to	19	17	4	3	13	12	2	2
82	this study is to	18	15	2	2	6	4	10	9
83	to account for the	15	11			14	10	1	1

84	to the fact that	15	15	4	4	5	5	6	6
85	total number of words	28	5			2	1	26	4
86	with regard to the	20	14	5	4	2	1	13	9
		2347	1381	647	394	661	410	1015	585

우선 두 코퍼스에서 나타난 어휘 묶음의 총 유형과 어휘수를 보면 EN의 경우 유형이 127개, 어휘수가 3087개로 나타났고 KN의 경우 85개, 어휘수가 2014개로 나타났다. 전체 빈도수의 합은 EN이 2505번, 텍스트의 합은 1572개로 나타났고 KN의 경우 각각 2347번, 1381개로 나타났다. 이 수치로는 두 코퍼스 간의 큰 차이는 보이지 않지만 반면에 학문별로 확인한 수치는 어느 정도의 차이를 보이고 있다. 먼저 EN의 경우 영문학에서 전체 빈도수의 합이 640개, 텍스트 수가 433개로 나타났고 영어학에서 각각 1079개, 666개, 영어교육학에서 각각 758, 473개로 나타났다. 영문학과 영어교육학 사이에는 큰 차이를 보이지 않지만 영어학에서 어휘 묶음이 훨씬 더 많이 나타나고 있는 것으로 확인되었다. 그러나 KN의 경우는 영문학에서 각각 647개, 394개, 영어학에서 661, 410개, 영어교육학에서 1015개, 585개로 나타나서 영문학과 영어학의 수치는 비슷한 반면 영어교육학에서 훨씬 많이 나타나고 있는 것으로 확인되었다. 나타난 어휘 묶음을 비교해 보면 ‘are more likely to’, ‘as a result of’, on the other hand, at the same time, the beginning of the, the fact that the, in terms of the, it is important to’ 등과 같이 공통적으로 많이 나타나는 것들도 보이고 반면에 ‘a great deal of, a wide range of, in a variety of, in the first place, in light of the’ 등은 EN에서만 15회 이상 나타나고 있고, ‘an important role in, different from that of, in spite of the, the results showed that, it is true that’ 등은 KN에서만 15회 이상 나타났다. 이제 두 코퍼스 간 빈도수가 많이 나타난 어휘 묶음에는 어떠한 것들이 있는지 상위 20개를 살펴보겠다. 그 결과

는 다음의 <표 10>과 같다.

<표 10> 두 코퍼스 간 어휘 묶음 상위 빈도 20개

N	English-Native		English-Non-native	
	어휘 묶음	빈도	어휘 묶음	빈도
1	at the same time	94	on the other hand	172
2	in the case of	83	at the same time	68
3	on the other hand	73	the end of the	64
4	the end of the	65	at the end of	63
5	the ways in which	62	in the case of	57
6	in the context of	57	in terms of the	54
7	in the united states	57	in the process of	54
8	as well as the	53	on the basis of	53
9	as a result of	45	the fact that the	52
10	it is important to	45	the results of the	51
11	on the basis of	43	one of the most	50
12	at the time of	40	as well as the	49
13	at the end of	39	as shown in table	46
14	on the one hand	39	the beginning of the	45
15	it is clear that	34	is one of the	44
16	the extent to which	34	in the form of	37
17	the use of the	34	at the beginning of	36
18	the fact that the	33	as a result of	31
19	in terms of the	31	in the sense that	31
20	in the form of	31	the purpose of this	28

결과를 보면 높은 빈도수로 나타난 어휘 묶음이기 때문에 같이 확인되는 것들이 많이 있다. 우선 각각 EN의 경우는 ‘at the same time’이 94번의 빈도로 가장 많이 나타났고 KN의 경우는 ‘on the other hand’가 172번의 빈도로 가장 많이 나타났다. EN에서는 73번의 빈도로 나타난 것으로 보면 KN에서 ‘on the other hand’의 사용이 두 배 이상의 차이로 훨씬 많이 사

용되어진 것으로 확인되었다. 그 외에 다른 것들은 빈도수에서 대체로 비슷하게 나타나면서 큰 차이는 보이고 있지 않았다. 이어서 ‘the end of the, in the case of, as well as the, as a result of, at the end of, the fact that the, in terms of the, on the basis of, in the form of’는 EN과 KN의 양쪽에서 모두 높은 빈도수로 나타났다. 반면에 ‘the ways in which, in the context of, it is important to, at the time of, on the one hand, it is clear that, the extent to which, the use of the’는 EN에서만 상위의 빈도수로 나타났고, ‘in the process of, the results of the, one of the most, as shown in figure, the beginning of the, is one of the, at the beginning of, in the sense that, the purpose of this’는 KN에서만 상위로 나타났다. ‘on the other hand’와 ‘on the one hand’ 같은 경우는 EN에서는 모두 상위에서 확인되고 있지만 KN에서는 ‘on the other hand’만 압도적으로 많이 사용되었지 전체 15회 이상의 빈도수 목록에서도 ‘on the one hand’는 나타나지 않는 것은 흥미 있는 결과라고 볼 수 있다. 다음으로 각 분야 당 많이 나타난 어휘 묶음을 살펴보겠다. 빈도수는 10회 이상 나타난 것으로 한다. 10회 이상 나타난 어휘 묶음 중에서 해당 분야에서만 많이 나타난 것에는 밑줄 그어 표시하였다. 그 결과는 다음의 <표 11>, <표 12>, <표 13>과 같다.

<표 11> 영문학 분야에서 많이 나타난 어휘 묶음

	English-Native	English-Non-native
어휘 묶음	<u>as a means of</u> as a result of <u>as part of a</u> as well as the at the end of	<u>and at the same</u> as a result of as well as the at the end of at the same time

<p>at the heart of at the same time in a way that in the case of in the context of in the face of in the first place in the form of in the united states it is clear that on the one hand on the other hand one of the most the beginning of the the end of the the fact that the the new york times the ways in which</p>	<p>can be seen as in sons and lovers in terms of the in the form of in the process of it is not the it is true that of the nineteenth century of the united states of the working class on the other hand one of the most the beginning of the the end of the the fact that the the house of the the structure of the the turn of the the ways in which</p>
---	---

먼저 영문학에서 많이 나타난 어휘 묶음에 대한 결과를 비교해 보면 위의 결과에서도 확인했지만 영문학 분야에서 나타난 어휘 묶음이 다른 분야에 비해 적기 때문에 영문학 분야에서만 많이 나타난 어휘 묶음도 그리 많지는 않은 것으로 확인되었다. EN의 경우 압도적으로 나타난 ‘the new york times’를 제외하고 ‘as a means of, as part of a, at the heart of, in the face of, in the first place’가 많이 나타났다. 반면 KN에서는 보편적으로 많이 사용되기 보다는 영문학이라는 주제에 한정하여 많이 나타나는 것으로 보여 지는 ‘in sons and lovers, of the nineteenth century, of

the working class, the house of the, the structure of the' 등이 나타났고 그 외에도 보편적으로 많이 사용될 수 있는 'and at the same, can be seen as, in the form of, it is not the, it is true that, the ways in which' 등이 많은 빈도수로 확인되었다.

<표 12> 영어학 분야에서 많이 나타난 어휘 묶음

	English-Native	English-Non-native
어휘 묶음	<u>a great deal of</u> <u>a result of the</u> <u>a wide range of</u> <u>are more likely to</u> as a result of <u>as shown in figure</u> <u>as we have seen</u> <u>as well as in</u> as well as the at the end of <u>at the level of</u> at the same time at the time of <u>can be seen in</u> <u>can be used to</u> <u>i would like to</u> <u>in a variety of</u> <u>in late middle english</u> <u>in relation to the</u> <u>in terms of the</u> in the case of in the context of	<u>argument of the verb</u> at the same time <u>in addition to the</u> <u>in spite of the</u> in terms of the in the case of <u>in the previous section</u> in the process of <u>in the same way</u> <u>in the sense that</u> is one of the native speakers of english <u>of this paper is</u> on the basis of on the other hand <u>that there is no</u> the fact that the <u>the presence of a</u> the purpose of this <u>the subject of the</u> <u>this paper is to</u> to account for the

	<p> <u>in the next section</u> <u>in the north and</u> <u>in the present study</u> <u>in the speech of</u> <u>in the united states</u> <u>in the use of</u> <u>is likely to be</u> <u>it is clear that</u> <u>it is difficult to</u> <u>it is important to</u> <u>it is possible that</u> <u>it is possible to</u> <u>it should be noted</u> <u>of the twentieth century</u> <u>on the basis of</u> <u>on the one hand</u> <u>on the other hand</u> <u>on the part of</u> <u>one of the most</u> <u>that there is a</u> <u>that there is no</u> <u>the beginning of the</u> <u>the case of the</u> <u>the end of the</u> <u>the fact that the</u> <u>the first half of</u> <u>the language of the</u> <u>the nature of the</u> <u>the rest of the</u> <u>the social meaning of</u> <u>the source of the</u> </p>	
--	--	--

	the use of the to account for the to the extent that	
--	--	--

다음으로 영어학 분야에서 많이 나타난 어휘 묶음에 대한 결과를 비교해 보면 나타난 종류에 있어서 EN이 KN에 비해 현저히 많은 것을 확인할 수 있었다. 이러한 결과를 토대로 확실히 단정 짓기는 무리가 있겠지만 EN의 경우 영어학에서 영문학이나 영어교육학에 비해 더 많은 종류의 어휘 묶음이 더 많은 빈도수로 나타나는 것으로 보아 주제의 특성상 정확한 데이터를 전달하기 위하여 미사여구 보다는 고정된 어휘 묶음을 사용한 간결한 문장을 더 선호한다고 볼 수도 있겠다. 확인된 어휘 묶음을 보면 EN의 경우 영어학이라는 주제로 인해 나타났다고 보여 지는 ‘in late middle english, in the speech of, of the twentieth century, the language of the, the social meaning of’ 등을 제외하고도 ‘a great deal of, a wide range of, are more likely to, can be used to, it is difficult to, it is possible that, on the part of, to account for the, to the extent that’ 등 다양한 종류의 어휘 묶음이 나타나는 것을 알 수 있었다. 반면에 KN에서는 ‘in addition to the, in spite of the, in the same way, in the sense that, the subject of the, this paper is to’ 등이 많은 나타나는 것으로 확인되었다. 이러한 결과를 바탕으로 우리나라의 영어 교육의 경우 어휘 중심의 교육이 많이 이루어지고 있고 더 나아가더라도 숙어나 관용어구 정도에서 그치고 있는 경우가 많은데 이러한 점을 보완하기 위해서는 좀 더 다양한 종류의 어휘 묶음을 사용할 수 있도록 이에 대한 교육이 더욱더 이루어져야 한다고 생각한다.

<표 13> 영어 교육학 분야에서 많이 나타난 어휘 묶음

	English-Native	English-Non-native
어휘 묶음	as a result of <u>as readers and writers</u> as well as the <u>at the beginning of</u> at the end of at the same time at the time of <u>in a way that</u> in the context of in the form of <u>in the process of</u> <u>in the role of</u> in the united states it is important to on the basis of <u>over the course of</u> <u>per cent of the</u> <u>teaching for social justice</u> the end of the <u>the extent to which</u> <u>the majority of the</u> <u>the new english education</u> <u>the time of the</u> the use of the the ways in which <u>ways in which the</u> <u>what it means to</u>	as a result of <u>as one of the</u> <u>as shown in table</u> as well as the <u>at the beginning of</u> at the end of at the same time <u>between the two groups</u> <u>can be said that</u> <u>difference between the two</u> <u>efl context of korea</u> <u>for the purpose of</u> in terms of the in the case of <u>in the efl context</u> <u>in the present study</u> in the process of <u>in the target language</u> <u>in the united states</u> <u>in the use of</u> is one of the <u>korean learners of english</u> native speakers of english <u>of the present study</u> <u>of the target language</u> <u>of this study is</u> on the basis of on the other hand

		one of the most the beginning of the <u>the efl context of</u> the end of the the fact that the <u>the purpose of the</u> the purpose of this <u>the result of the</u> <u>the results of the</u> <u>the results showed that</u> <u>the total number of</u> <u>the two types of</u> <u>the use of the</u> <u>this study is to</u> <u>total number of words</u> <u>with regard to the</u>
--	--	--

마지막으로 영어교육학 분야에서 많이 나타난 어휘 묶음에 대한 결과를 비교해 보면 위의 영어학 분야에서와는 반대로 EN보다 KN에서 더 많은 종류의 어휘 묶음이 나타나는 것으로 확인되었다. EN의 경우 ‘at the beginning of, in a way that, in the process of, in the role of, the extent to which, what it means to’ 등이 많이 나타났고 KN의 경우 ‘as one of the, at the beginning of, can be said that, for the purpose of, the results of the, the use of the, this study is to, with regard to the’등이 나타났다. 특히 KN의 경우 ‘efl context of korea, in the target language, korean learners of english, the efl context of, the total number of, total number of words’ 등은 영어교육학 쪽에서만 편중되어 집중적으로 많이 나타난 것으로 확인되었다.

4.3.2 어휘 묶음의 구조에 따른 사용 비교 분석

앞서 두 코퍼스에서 확인된 다양한 어휘 묶음의 종류와 빈도, 학문별로 많이 나타난 어휘 묶음을 살펴보았다. 본 논문에서는 다음으로 어휘 묶음의 구조에 따라 분류해 보고 그 양상을 확인해보았다. 분류는 Cortes(2004)의 방법을 이용하였다. 그에 따른 결과는 다음의 <표 14>, <표 15>와 같다.

<표 14> English-Native의 어휘 묶음의 구조에 따른 분류

<p>1. Noun phrase with 'of' phrase fragment</p> <p>a great deal of, a result of the, a wide range of, one of the most, per cent of the, the beginning of the, the case of the, the context of the, the end of the, the first half of, the history of the, the language f the, the majority of the, the nature of the, the part of the, the rest of the, the role of the, the social meaning of, the source of the, the time of the, the use of the</p>
<p>2. Noun phrase with post-nominal clause fragment</p> <p>the extent to which, the fact that the, the way in which, the ways in which, ways in which the</p>
<p>3. Prepositional phrase with embedded 'of' phrase</p> <p>at the beginning of, at the end of, at the heart of, at the level of, at the time of, in a variety of, in light of the, in terms of the, in the case of, in the context of, in the development of, in the face of, in the form of, in the process of, in the role of, in the speech of, in the use of, on the basis of, on the part of, over the course of</p>
<p>4. Other prepositional fragments</p> <p>as a means of, as a result of, as part of a, as part of the, at the same time, despite the fact that, in a way that, in other words the, in relation to the, in response to the, in the first place, in the next section, in the present study, in the same way, in the united states, of the relationship between, of the united states, on the one hand, on the other hand, to</p>

account for the, to the extent that
5. Verb (be) + complement (noun phrase)
is one of the,
6. (modal) Vbe+(complement noun phrase/adj. phrase)
are more likely to, can be seen in, can be used to, is likely to be
7. it+Vbe+adjective+(clause fragment)
it is clear that, it is difficult to, it is important to, it is necessary to, it is not the, it is possible that, it is possible to
8. Other expressions
as readers and writers, as shown in figure, as we have seen, as well as in, as well as the, I would like to, it should be noted, that there is a, that there is no, to do with the, what it means to

<표 15> English-Non-native의 어휘 묶음의 구조에 따른 분류

1. Noun phrase with 'of' phrase fragment
one of the most, the basis of the, the beginning of the, the content of the, the end of the, the meaning of the, the presence of a, the purpose of the, the purpose of this, the result of the, the results of the, the structure of the, the subject of the, the total number of, the turn of the, the two types of, the use of the, total number of words
2. Noun phrase with post-nominal clause fragment
the fact that the, the ways in which
3. Other Noun Phrase
an important role in, difference between the two, the relationship between the,
4. Prepositional phrase with embedded 'of' phrase
at the beginning of, at the end of, for the purpose of, in terms of the, in the case of, in the context of, in the form of, in the process of, in the use of,
5. Other prepositional fragments
as a result of, as one of the, at the same time, between the two groups, in addition to the, in relation to the, in spite of the, in the present study, in the previous section, in the same way, in the sense that, in the united states, of the present study, of this paper is, of this

study is, on the basis of, on the other hand, to account for the, to the fact that, with regard to the
6. Verb (be) + complement (noun phrase)
be one of the, is one of the,
7. (modal) Vbe+(complement noun phrase/adj. phrase)
are more likely to, can be said that, can be seen as, does not seem to, is based on the,
8. it+Vbe+adjective+(clause fragment)
it is important to, it is necessary to, it is not the, it is true that,
9. (noun phrase/pronoun)+V+(complement)
the results showed that, this paper is to, this study is to
10. Other expressions
and at the same, as shown in table, as shown in the, as well as the, different from that of, it can be said, that there is no,

결과를 보면 전체적으로 EN과 KN 모두 ‘the beginning of the, one of the most’와 같이 전치사 ‘of’가 포함되어 있는 명사구와 ‘in terms of the, in the case of’와 같은 ‘of’가 포함되어 있는 전치사구의 구조가 많이 나타나는 것을 확인할 수 있었다. 이 점은 Cortes(2004)의 연구 결과와도 같다고 할 수 있다. 다음으로 뒤에서 꾸며주는 절이 포함된 명사구의 경우 EN과 KN 모두 ‘the fact that the, the ways in which’가 확인되었고 ‘the extent to which’의 경우는 EN에서만 나타났다. 그 외에 다른 명사구로 이루어진 어휘 묶음의 경우는 EN에서는 나타나지 않았고 KN에서만 ‘an important role in, difference between the two, the relationship between the’ 이렇게 세 종류가 나타났다. 또한 ‘of’가 포함된 명사구와 전치사구 외에도 다른 전치사구 역시 많이 확인되었다. EN에서는 ‘despite the fact that, in other words the, in response to the, in the same way, to the extent that’ 등이 확인되었고 KN에서는 ‘in addition to the, in relation to the, in spite of the, on the basis of, with regard to the’등이 확인되었다. 다음으로 동사 다음에 명사구

인 보어로 이루어진 구조로는 EN과 KN 모두 'is(be) one of the'만 확인되었다. 다음으로 서법 조동사나 be동사와 뒤이어 나오는 명사구나 형용사구의 보어로 이루어진 구조로는 EN에서는 'are more likely to, can be seen in, can be used to, is likely to be'가 확인되었고 KN에서는 'are more likely to, can be said that, can be seen as, does not seem to, be based on the'가 확인되었다. 다음으로 it을 주어로 뒤이어 be동사와 형용사, 그리고 절로 이어지는 구조를 가진 어휘 묶음에는 'it is important to, it is necessary to, it is not the'는 EN과 KN에서 공통으로 나타났고 'it is clear that, it is difficult to, it is possible that(to)'는 EN에서만, 'it is true that'은 KN에서만 나타났다. 그리고 'of'가 없는 명사구 외에도 EN에는 없지만 KN에서만 나타나는 구조가 하나 더 있었다. 명사 혹은 대명사를 주어로 뒤이어 동사와 보어가 있는 구조인데 그 예로는 'the results showed that, this paper is to, this study is to'등이다. 그 외에 EN에서는 'as we have seen, I would like to, it should be noted, what it means to'등이 확인되었고 KN에서는 'as shown in table, different from that of, it can be said, that there is no'등이 확인되었다.

4.3.3 어휘 묶음의 기능에 따른 사용 비교 분석

마지막으로 두 코퍼스에 나타난 어휘 묶음을 기능에 따라서 분류해 보고 그 사용 양상을 확인해 보았다. 기능에 따른 분류 역시 Cortes(2004)의 방법을 이용하였다. 그에 따른 결과는 다음의 <표 16>, <표 17>과 같다.

<표 16> English-Native의 어휘 묶음의 기능에 따른 분류

범주	하위 범주	어휘 묶음(개수)
지시 어휘 묶음 (Referential bundles)	시간 어휘 묶음 (Time markers)	at the beginning of, at the end of, at the same time, at the time of, in late middle english, in the development of, in the process of, of the twentieth century, over the course of, the beginning of the, the end of the, the time of the (12)
	장소 어휘 묶음 (Place markers)	in the first place, in the next section, in the north and, in the present study, in the speech of, in the united states, of the relationship between, of the united states (8)
	묘사 어휘 묶음 (Descriptive bundles)	as readers and writers, of the white elephant, teaching for social justice, the history of the, the language of the, the nature of the, the new english education, the new york times, the content of the (9)
	양화 어휘 묶음 (Quantifying bundles)	a great deal of, as part of a, as part of the, in a variety of, is one of the, on the part of, one of the most, per cent of the, that there is a, that there is no, the extent to which, the

		first half of, the majority of the, the part of the, the rest of the, to the extent that (16)
텍스트 구성 어휘 묶음 (Text Organizer)	대조/비교/추론 (Contrast/Comparison Inferential)	a result of the, as a result of, in relation to the, on the basis of, on the one hand, on the other hand, the social meaning of, the source of the (8)
	초점(Focus)	it is difficult to, it is important to, it is necessary to, it should be noted (4)
	틀형성(Framing)	as well as in, as well as the, despite the fact that, in a way that, in light of the, in other words the, in response to the, in terms of the, in the case of, in the context of, in the face of, in the form of, in the role of, in the same way, in the use of, the case of the, the context of the, the fact that the, the role of the, the use of the, the way in which, the ways in which, ways in which the (23)
견지 어휘 묶음 (Stance bundles)	인식 어휘 묶음 (epistemic-impersonal/ probable-possible)	are more likely to, can be seen in, is likely to be, it is clear that, it is not the, it is possible that, it is possible to (7)
	기타 인식 어휘 묶음 (Other stance bundles)	as shown in figure, as we have seen (3)

기타 어휘 묶음 (Other bundles)		a wide range of, as a means of, at the heart of, at the level of, can be used to, I would like to, to account for the, to do with the, what it means to (9)
-----------------------------	--	---

<표 17> English-Non-native의 어휘 묶음의 기능에 따른 분류

범주	하위 범주	어휘 묶음(개수)
지시 어휘 묶음 (Referential bundles)	시간 어휘 묶음 (Time markers)	and at the same, at the beginning of, at the end of, at the same time, in the process of, of the nineteenth century, the beginning of the, the end of the (8)
	장소 어휘 묶음 (Place markers)	between the two groups, in the present study, in the previous section, in the target language, in the united states, of the present study, of this paper is, of this study is (8)
	묘사 어휘 묶음 (Descriptive bundles)	argument of the verb, difference between the two, efl context of korea, in sons and lovers, korean learners of english, native speakers of english, of the target language, of the united states, of the working class, the house of the, the purpose of this, the structure of the, the turn of the (13)

	양화 어휘 묶음 (Quantifying bundles)	as one of the, be one of the, is one of the, one of the most, that there is no, the total number of, total number of words (7)
텍스트 구성 어휘 묶음 (Text Organizer)	대조/비교/추론 (Contrast/Comparison Inferential)	as a result of, in relation to the, in spite of the, is based of the, on the basis of, on the other hand, the basis of the, the meaning of the, the result of the, the results of the, the results showed that (11)
	초점(Focus)	an important role in, it is important to, it is necessary to (3)
	틀형성(Framing)	as well as the, for the purpose of, in addition to the, in terms of the, in the case of, in the context of, in the efl context, in the form of, in the same way, in the sense that, in the use of, the efl context of, the fact that the, the presence of a, the subject of the, the use of the, the ways in which, this paper is to, this study is to, to the fact that, with regard to the (21)
견지 어휘 묶음 (Stance)	인식 어휘 묶음 (epistemic-impersonal/ probable-possible)	are more likely to, can be said that, can be seen as, different from that of, does not seem to, it

		can be said, it is not the, it is true that (8)
bundles)	기타 인식 어휘 묶음 (Other stance bundles)	as shown in table, as shown in the (2)
기타 어휘 묶음 (Other bundles)		the relationship between the, to account for the (2)

결과를 보면 전체적으로 큰 범주 안에서는 지시 어휘 묶음 많이 나타나고 있다. 뒤이어 텍스트 구성 어휘 묶음과 견지 어휘 묶음이 나타나는데 세부적 범주로 다시 나누어서 보면 시간 어휘 묶음으로는 EN에서 KN보다 조금 더 많이 나타나는 것으로 확인되었다. ‘at the same time, at the beginning of, at the end of, in the process of’등은 공통으로 나타나는 반면 ‘at the time of, in the development of, over the course of’등은 EN에서만 나타났고 KN에서는 KN에서만 나타나는 눈에 띄는 어휘 묶음은 확인되지 않았다. 다음으로 장소 어휘 묶음을 보면 ‘in the united states’만 공통으로 나타나는 것 외에는 EN에서는 ‘in the first place, in the next section, of the relationship between’등이 확인되었고 KN에서는 ‘in the present study, in the previous section, of this paper is, of this study is’등이 확인되었다. 다음으로 묘사 어휘 묶음은 논문을 쓰면서 해당 내용을 제시하기 위해 사용되기 때문에 공통적으로 나타나는 것은 없었다. 각각 EN에서는 ‘the history of the, the nature of the, the content of the’등이 나타났고 KN에서는 ‘the purpose of this, the structure of the, the turn of the’등이 나타났다. 다음으로 quantifying bundles는 EN에서 KN보다 좀

더 여러 종류의 어휘 묶음이 확인되었는데 EN에서는 ‘a great deal of, as part of the, in a variety of, is one of the, one of the most, the first half of, the majority of the, the rest of the, that there is no, to the extent that’등이 확인되었고 KN에서는 ‘as one of the, is one of the, one of the most, that there is no’등 EN에서도 나타난 어휘 묶음만 공통적으로 확인되었다.

다음으로 큰 범주 텍스트 구성 어휘 묶음을 살펴보면 전반적으로 두 코퍼스에서 모두 틀형성의 기능이 많이 나타났다. 세부적 범주로 나누어서 보면 먼저 대조/비교/추론의 기능을 하는 어휘 묶음으로는 EN에서는 ‘as a result of, in relation to the, on the basis of, on the one hand, on the other hand’ 등이 나타났고 KN에서는 ‘as a result of, in relation to the, on the basis of, on the other hand’등은 공통으로 나타났고 그 외에 ‘in spite of the, is based on the, the meaning of the, the results showed that’등이 나타났다. 다음으로 초점 기능을 하는 어휘 묶음으로는 두 코퍼스 모두에서 it is important to, it is necessary to’가 나타났고 그 외에 각각 EN에서는 ‘it is difficult to, it should be noted’가, KN에서는 ‘an important role in’이 나타났다. 다음으로 가장 많이 나타난 틀형성 기능을 살펴보면 ‘as well as the, in terms of the, in the case of, in the context of, in the form of, in the same way, in the use of, the fact that the, the ways in which’등은 공통으로 나타났고 ‘in light of the, in other words the, in response to the, in the face of, the role of the’는 EN에서만 나타났고 ‘for the purpose of, in addition to the, in the sense that, the subject of the, this paper is to, this study is to, with regard to the’등은 KN에서만 나타났다.

마지막으로 큰 범주인 견지 어휘 묶음을 살펴보면 하위 범주인 인식 어휘

묶음으로는 ‘are more likely to’와 ‘it is not the’가 공통으로 나타났고 각각 EN에서는 ‘can be seen in, it is clear that, it is possible that(to)’ 등이 나타났고 KN에서는 ‘can be said that, can be seen as, different from that of, does not seem to, it is true that’등이 나타났다. 그 외에 다른 어휘 묶음으로는 EN에서는 ‘a wide range of, can be used to, to account for the, to do with the’등이 있었고 KN에서는 ‘the relationship between the, to account for the’ 등이 있었다.

위의 결과를 바탕으로 보면 지시 어휘 묶음에 있어 하위 범주의 어휘 묶음들의 수에서 다른 것들은 큰 차이가 없으나 양화 어휘 묶음에 있어서 EN이 16개가 나타났고 반면에 KN에서는 7개가 나타났다. 양화 어휘 묶음 사용에 대한 교육이 좀 더 이루어져야 할 것으로 보인다. 또한 확인된 기타 어휘 묶음도 EN이 9개, KN이 2개로 EN이 좀 더 다양하게 나타났다. 이러한 어휘 묶음 사용에 대한 교육에 있어서 좀 더 다양성을 갖출 필요가 있을 것이다. 그리고 어휘 묶음의 세부적 기능 중에서 EN과 KN에서 각각 20개가 넘는 틀형성 기능을 하는 어휘 묶음이 가장 많이 확인된 것으로 보아 학술 논문이나 학문적 글쓰기에 있어서 많이 사용되고 필요한 만큼 이 부분에 대한 교육 역시 중요하게 다루어져야 할 것이다.

4.4 접속(부)사 사용의 비교 분석

매끄럽고 자연스러우며 논리적인 글을 쓰기 위해서 결속성은 상당히 중요한 요소로 작용한다. 글을 쓰는데 있어서 주요한 어휘나 내용을 달라지게 만들 수 있는 문법적인 요소들이 중요하기 때문에 다소 글의 결속성에 대해서는 소홀해 질 수 있다. 하지만 제대로 된 글의 완성에는 영어의 접속사나 접속부사를 적절히 사용하여 얼마나 결속력 있게 쓰느냐에 달려 있다고 할 수

있다. 이러한 측면에 있어서 원어민과 비원어민 간의 접속사나 접속부사 사용 양상에 대한 비교 연구도 다양하게 이루어지고 있는데 본 논문에서도 수집한 두 코퍼스 사이에 나타난 접속사와 접속부사의 사용을 통해 기존의 연구 결과와도 비교 분석하면서 그 사용 양상을 알아보려 한다. 먼저 각각 EN과 KN에서 접속사와 접속부사가 얼마만큼의 빈도로 나타나는지 확인해보고 접속(부)사의 경우는 사용할 수 있는 위치가 한 곳으로 정해져 있지 않기 때문에 그에 따른 차이점도 있는지 나타나는 위치 또한 함께 알아보았다. 해당 결과는 다음의 <표 18>과 같다.

<표 18> 접속(부)사의 빈도수 & 위치

	English-Native					English-Non-native				
	빈도	위치				빈도	위치			
		문두	비율	문중	비율		문두	비율	문중	비율
however	1058	547	51.7	511	48.3	1033	720	69.7	313	30.3
but	3113	463	14.9	2650	85.1	1883	229	12.2	1654	87.8
and	27776	426	1.5	27350	98.5	21447	260	1.2	21187	98.8
although	592	353	59.6	239	40.4	344	212	61.6	132	38.4
when	1403	293	20.9	1110	79.1	1264	359	28.4	905	71.6
because	929	100	10.8	829	89.2	933	67	7.2	866	92.8
while	1036	355	34.3	681	65.7	640	151	23.6	489	76.4
so	1460	153	10.5	1307	89.5	1012	151	14.9	861	85.1
yet	390	123	31.5	267	68.5	185	63	34.1	122	65.9
therefore	368	64	17.4	304	82.6	399	280	70.2	119	29.8
thus	512	176	34.4	336	65.6	523	296	56.6	227	43.4

결과를 보면 먼저 역접의 의미를 갖고 있는 ‘however’과 ‘but’을 보면 두 접속(부)사 모두 일반적으로 글을 쓰는데 있어서 상당히 많이 나타난다. 빈도의 측면에서 보면 ‘however’는 큰 차이를 보이지 않지만 ‘but’의 사용은 EN에서 훨씬 많이 나타나고 있는 것을 확인할 수 있다. 사용 위치 측면에서

보면 'but'과 같은 경우는 EN과 KN에서 각각 문두와 문중에 나타난 비율이 EN의 경우 문두가 14.9, 문중이 85.1로 나타났고 KN의 경우 문두가 12.2, 문중이 87.8로 비슷하게 나타나면서 두 자료 모두 but이 문두보다는 문중에서 많이 사용되고 있는 것으로 확인되었다. 반면에 'however'의 경우 EN과 KN간의 사용 위치에 있어서 차이를 보였다. 각각 EN의 경우 문두가 51.7, 문중이 48.3으로 비슷하게 나타나고 있는 반면에 KN의 경우 문두가 69.7, 문중이 30.3으로 문두에서의 사용이 문중보다 두 배 이상 차이가 나는 것으로 확인되었다. 원어민에 비해 한국인 화자의 글에서 'however'의 사용이 문두로 집중되는 것은 사용 양상에 있어서 큰 차이임은 분명하다고 본다. 'however'의 사용은 많은 글에서 높은 빈도를 보이는 어휘인 만큼 다른 연구들에서도 볼 수 있는데 고광윤(2008)은 영어교과서와 미국영어를 바탕으로 한 Frown 코퍼스 사이에 사용 양상을 비교하였는데 이 연구에서는 Frown 코퍼스에서는 'however'의 사용이 문두보다 문중에 쓰이는 경우가 압도적으로 많이 나온 반면 영어 교과서에서는 문두에서의 사용이 훨씬 많음을 확인하였다. 또한 영어 교과서뿐만 아니라 대학생들의 영작문이나 영자신문에서도 비슷한 결과를 나타냈다고 제시하였다. 본 논문의 결과와 비교해보았을 때 원어민 코퍼스의 결과에서는 다소 차이가 있었지만 한국인 화자의 코퍼스에서는 같은 결과를 보였다. 영어를 전공하는 한국인 화자의 학술논문에서도 이러한 결과가 나온 것으로 미루어 고광윤(2008)이 밝힌 대로 영어 교과서에서 'however'가 문두에 훨씬 더 많이 나타나고 있기 때문에 한국인 학습자가 영어 교육을 받아오기를 'however'를 문두에 쓰는 경우를 더 많이 보고 학습할 수밖에 없고 그로 인해 자연스럽게 한국어 화자들이 글을 쓰는데 있어 'however'를 문두에 사용하는 것을 더 선호하는 경향이 있는 것으로 생각해 볼 수 있다. 이런 결과를 생각해 볼 때 영어 교육에 있어 한쪽으로 편중된 위치만을 학습시키기보다 다양한 가능성을 함께 제시할

필요가 있을 것이다.

다음으로 상당히 많은 빈도수로 나타나는 접속사 ‘and’를 비교해 보면 빈도수로 는 두 코퍼스 모두에서 많이 나타나고 있는데 사용 위치의 측면에서 볼 때 EN과 KN모두 문종에서 압도적으로 많이 사용되고 있다. 일반적으로 ‘and’는 문두에서 잘 사용하지 않는 것으로 학습해 오지만 각각 1.5, 1.2의 비율로 나타나기도 하는 것을 확인할 수 있다. 다음으로 많이 쓰이는 접속사 ‘although, when, because, while’을 살펴보면 빈도수에서는 ‘while’만 1036번, 640번으로 차이가 약간 있고 그 외에는 모두 비슷하게 나타났다. 사용 위치 측면에서도 큰 차이 없이 ‘although’는 두 코퍼스 모두 문두에서 많이 나타났고, ‘when’과 ‘because’, ‘while’의 경우는 문종이 압도적으로 많이 나타났다.

다음으로 ‘so’의 경우를 살펴보면 각각 두 코퍼스에서 빈도수 1460, 1012번으로 나타나 높은 빈도수를 보이고 있다. 또한 사용 위치를 비교해 보면 두 코퍼스에서 모두 문종에서 압도적으로 많이 사용하고 있는 것으로 확인되었다. 구선화(2012)는 원어민과 한국인 영어 학습자의 학술지 논문에서 나타난 접속부사의 사용 양상을 비교하였는데 한국인의 학술지 논문에서 ‘so’의 사용이 두 배가 많은 것으로 한국인이 ‘so’를 자주 사용한다고 하였는데 본 논문에서의 결과에서는 그와는 반대로 원어민이 큰 차이는 아니지만 더 많이 사용하는 것으로 나왔다. 이러한 경향을 좀 더 확실하게 확인하기 위해서는 좀 더 다양한 글을 통한 폭넓은 연구가 이루어져야 할 것이다.

다음으로 ‘yet’의 경우 빈도수에서는 EN에서 좀 더 많이 나타나지만 사용 위치에서는 두 코퍼스 모두 문종에서 많이 나타나고 있었다.

마지막으로 비슷한 의미를 갖고 있는 ‘therefore’과 ‘thus’를 비교해 보겠다. 먼저 빈도수의 측면에서 보면 두 코퍼스에서 모두 ‘therefore’에 비해서 ‘thus’가 많이 사용되고 있는 것으로 나타났다. 하지만 EN과 KN간에 빈도

수의 차이는 거의 나지 않았다. EN과 KN에서 모두 비슷하게 사용되고 있다고 할 수 있다. 이러한 본 논문의 결과를 역시 다른 결과와 비교해 보면 구선화(2012)는 원어민과 한국인 화자 간에 ‘therefore’에 있어서는 비슷하지만 ‘thus’의 경우는 원어민 화자가 더 많이 사용한다고 제시하여 본 논문과는 약간 다른 결과라고 할 수 있다. 이 역시도 좀 더 확실한 사용양상을 확인하기 위해서는 추가 연구가 더욱 이루어져야 할 것이다. 이어서 문장 위치 측면에서 살펴보면 두 코퍼스 간에 차이가 있음을 알 수 있다. ‘therefore’ 경우 EN에서는 문두가 17.4, 문중이 82.6으로 문중에서 압도적으로 많이 사용되고 있는 반면에 KN에서는 문두가 70.2, 문중이 29.8로 문두에서 압도적으로 많이 사용되고 있었다. 본 논문의 이러한 결과는 고헌윤(2008)의 연구 결과와 일맥상통한다. 이 연구에서도 ‘therefore’의 사용이 현대 미국영어에서 문중에서 사용하는 것이 대부분이고 영어 교과서에서는 문두가 압도적으로 많음을 제시하였다. 또한 ‘thus’의 경우 EN에서는 문두가 34.4, 문중이 65.6으로 문중에서 좀 더 많이 사용되었고 KN에서는 문두가 56.6, 문중이 43.4로 문두에서 좀 더 많이 사용되었다. 이러한 결과들을 토대로 원어민 화자의 경우 접속부사를 문두, 문중 위치에 다양하게 쓰되 문중에서 더 많이 사용하고 있음을 확인했으나 한국인 화자의 경우 접속부사 사용에 있어서 문중보다는 문두에 편중되는 경향이 강한 것으로 확인되었다. 이렇듯 다양한 종류의 글에서 접속부사의 사용 위치에 대한 차이가 남으로써 이를 해소하기 위해서는 영어 교과서에서도 편중된 위치 사용이 아닌 다양한 사용의 용례를 보임으로써 학습자가 자연스럽게 익힐 수 있도록 할 필요가 있고 원어민이 작성한 좀 더 다양한 글을 접할 수 있는 기회를 많이 마련해줘야 할 것이다.

4.5 참조 코퍼스에 따른 중요 어휘 비교 분석

4.5.1 BNC 참조 중요 어휘 비교 분석

본 논문에서 앞서 두 코퍼스 각각의 어휘 목록의 상위에 나타난 내용어와 기능어를 나누어 그 사용 빈도수를 참조하여 비교해보았다. 다음으로 일반적으로 가장 많이 사용되고 있는 BNC 코퍼스를 참조하여 각각의 코퍼스에서 두드러지게 나타나는 중요 어휘를 알아보았다. 코퍼스 프로그램 WordSmith Tools 5.0을 사용하여 추출한 중요 어휘 목록에서 참조 코퍼스에 비해 얼마나 많이 나타났는지를 알려주는 중요도(keyness)를 중점으로 수치 1000 이상에서 나타난 중요 어휘를 살펴보았다. 그 비교 목록은 다음의 표와 같다.

<표 19> BNC 참조 두 코퍼스 간의 중요 어휘 목록 비교

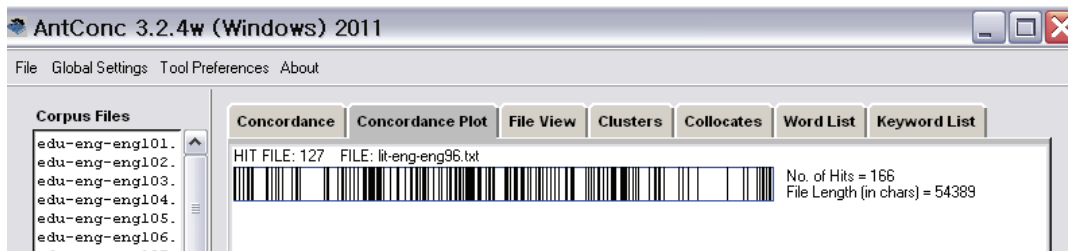
N	English-Native		English-Non-native	
	중요 어휘	중요도	중요 어휘	중요도
1	Students	9117	English	11926
2	English	6739	Learners	11468
3	Teachers	6282	Korean	10283
4	Language	4316	Students	7236
5	Teacher	4108	EFL	5577
6	Learners	3778	Verb	5182
7	Preservice	3590	Language	4869
8	Speakers	3295	Feedback	3919
9	Linguistic	3139	Verbs	3796
10	Classroom	3027	Proficiency	3775
11	Writing	2794	Learning	3402
12	Poetry	2565	Self	3239
13	Dialect	2477	Constructions	3112
14	As	2477	VP	2995
15	Dickens	2451	Vowel	2955
16	Teaching	2450	Writing	2835

17	Self	2290	Speakers	2748
18	Reading	2249	NP	2653
19	Contexts	2123	Native	2502
20	Verb	2118	Whitman	2484
21	Of	1977	Study	2479
22	Student	1945	Non	2358
23	Texts	1929	Participants	2279
24	In	1880	Reading	2138
25	Epistemic	1847	Sentences	2098
26	Participants	1821	Words	2081
27	Frequency	1799	In	2080
28	Variation	1767	Syntactic	2075
29	Meaning	1752	Sentence	1903
30	Learning	1746	Vowels	1780
31	Data	1733	Learner	1760
32	Verbs	1714	Clauses	1653
33	Fricatives	1643	Comprehension	1650
34	Lexical	1629	Adjunct	1642
35	Lame	1594	Is	1633
36	Study	1538	Syllable	1609
37	Class	1530	Analysis	1606
38	Social	1482	As	1579
39	Percent	1462	Of	1523
40	Novel	1443	Teacher	1518

결과를 보면 보편적으로 많이 사용되는 어휘인 BNC와 비교하였기 때문에 각각 EN에서는 ‘students’가 중요도 9117로, KN에서는 ‘English’가 중요도 11926으로 가장 두드러지는 어휘로 나타났다. 그 아래로는 두 코퍼스 간에 비슷하게 ‘learners, language, writing’등의 어휘가 확인되었다. 차이점을 비교하자면 EN의 경우 ‘preservice’나 ‘classroom’등이 나타나는 반면에 KN에서는 아무래도 한국인 화자의 학술 논문이다 보니 ‘Korean, EFL’등의 어휘가 중요도가 높게 나타나고 있다. 또한 EN의 경우 KN에서 나타나지 않는 ‘dialect, variation’ 등의 어휘가 나타남으로써 최근의 연구 경향이 사투리나 변이에 대한 것이 많이 이루어지고 있음을 생각해 볼 수 있다. 또한

‘texts, frequency’등의 어휘로 이루어진 코퍼스 언어학에 대한 연구 또한 많이 이루어지고 있음을 예측해 볼 수 있다. 반면에 KN의 경우 한국인의 경우 최대한 원어민에 가까운 영어 구사를 목표로 학습하는 환경에서 아무래도 그에 관한 연구들이 많이 이루어지기 때문에 ‘feedback, proficiency, native’와 같은 어휘들이 중요하게 나타나고 있다. 그 외에 ‘sentence, syntactic, constructions, clauses, adjunct, VP, NP’와 같이 영어의 구문론적인 특징을 나타내는 어휘 또한 두드러지게 나타나고 있다. 그 외에도 각각의 코퍼스에서 나타나고 있는 이름으로 EN의 경우 ‘Dickens’만이 나타나고 있고 KN의 경우 ‘Whitman’만 나타나고 있다. 한편 조심해야 할 것이 있다. 또한 주로 내용어 중 명사가 두드러지게 나타나는 와중에 확인되는 기능어들이 나타났는데 접속사 ‘as’와 전치사 ‘of, in’이 중요도에서는 차이가 있지만 두 코퍼스에서 공통으로 확인되었다. 접속사 ‘as’의 경우 다양한 뜻으로 사용될 수 있기 때문에 인과관계가 중요시 되는 학술논문에서 많이 나타나고 있다고 여겨지고 전치사 ‘of’의 경우 학술논문에서 명사구의 사용이 많음을 반영한다고 볼 수 있다. 전치사 ‘in’의 경우도 학술 논문의 특성상 어떤 자료나 데이터 안에서 결과를 제시해야 하기 때문에 이를 나타내기 위해 많이 사용되는 것으로 볼 수 있겠다. 마지막으로 중요 어휘를 확인할 시에 주의해야 할 점이 있는데 이는 EN의 목록에서 보면 본문의 표에는 실려있지 않지만 ‘tarzan’이라는 어휘가 중요도 1251로 확인되며 상위에 나타나고 있는데 이 어휘가 유의미 하려면 수집된 자료에 다양하게 등장해야만 한다. 이를 확인하기 위해 Antconc 프로그램의 concordance plot 이라는 기능을 사용한다. 이 기능은 해당 어휘가 텍스트에서 어떠한 분포를 보이는지 나타내준다. 해당 어휘 ‘tarzan’을 확인해 본 결과는 다음의 <그림 1>과 같다.

<그림 1> concordance plot을 이용한 ‘tarzan’의 분포



위의 그림처럼 ‘tarzan’의 분포가 한 텍스트에서만 166번이 나타난 것을 확인할 수 있다. 이처럼 코퍼스를 이용하여 데이터를 분석할 때 무조건 그 수치만을 믿고 생각할 것이 아니라 그 분포에 있어서 이 어휘가 중요 어휘로써 유의미한지를 알기 위해서 정확한 확인과 주의가 필요할 것이다.

4.5.2 원어민 화자와 한국인 화자의 코퍼스 참조 중요 어휘 비교 분석

위에서 각각의 코퍼스를 BNC 코퍼스와 참조하여 확인된 중요 어휘들에는 어떠한 차이가 있는지 비교해 보았다. 다음으로 서로의 코퍼스를 참조하여 중요 어휘를 확인해 보았다. 이것은 서로의 코퍼스를 참조하여 나온 결과이기 때문에 각각 원어민 화자의 영어 관련 학술 논문과 한국인 화자의 영어 관련 학술 논문에서 더 두드러지게 나타나는 어휘가 무엇인지 비교해 볼 수 있다. 따라서 두 코퍼스 간의 중요 어휘간의 차이를 비교해 보는데 있어서 더 용이하다고 할 수 있다. 역시 위의 방법과 마찬가지로 코퍼스 프로그램 WordSmith Tools 5.0을 사용하여 추출한 중요 어휘 목록에서 참조 코퍼스에 비해 얼마나 많이 나타났는지를 알려주는 중요도(keyness)를 중점으로 100 이상에서 나타난 중요 어휘를 살펴보았다. 그 비교 목록은 다음의 표와 같다.

<표 20> EN과 KN의 코퍼스 참조 중요 어휘 목록 비교

N	English-Native		English-Non-native	
	중요 어휘	중요도	중요 어휘	중요도
1	We	578	Korean	2255
2	Dickens	480	The	617
3	Preservice	456	Learners	589
4	Social	427	Feedback	586
5	These	415	EFL	551
6	Jones	383	Proficiency	439
7	Teachers	367	Sentence	435
8	Dialect	349	Korea	414
9	Our	323	Native	386
10	Lame	276	English	367
11	Percent	253	Sentences	344
12	Would	211	VP	344
13	Newfoundland	204	John	342
14	Variation	197	Adjunct	309
15	Tarzan	191	NP	288
16	Rivers	184	Whitman	285
17	Anita	176	His	284
18	Brandy	172	Test	282
19	Nuclear	171	Comprehension	264
20	Burne	170	Structure	234
21	Labov	169	Predicate	223
22	Work	169	Bradley	223
23	Frequency	165	Readings	219
24	Bishop	163	Target	217
25	Data	163	Verb	217
26	Often	157	Object	205

27	Adolescents	156	Margery	204
28	Northern	151	Lacan	204
29	Variant	150	Maximal	202
30	Dialects	147	Ellipsis	196

결과를 비교해 보면 우선 위의 BNC 코퍼스를 참조로 한 중요 어휘 목록과 확연한 차이를 알 수 있다. EN의 경우 특이하게도 인칭 대명사 ‘we’가 중요도 578로 KN에 비해 가장 두드러지게 많이 나타나는 어휘로 확인되었다. 이것은 본 논문의 대명사 빈도 연구에서 KN에 비해 EN에서 ‘we’가 두 배 이상 많이 사용되고 있음을 확인한 결과와 일맥상통한다고 볼 수 있다. 반면에 KN에서는 위의 결과와 비슷하게 ‘Korean’이 중요도 2255의 압도적인 수치로 중요 어휘로 확인되었다. 뒤이어 EN에서는 ‘Dickens, preservice, social, these’ 등의 어휘가 중요 어휘로 나타났으며 KN에서는 ‘the, learners, feedback, EFL’등이 나타났다. 주목할 만한 차이점으로는 EN에서 대명사 ‘these’와 서법조동사 ‘would’가 목록에서 확인되고 있는데 이 역시도 본 논문의 대명사와 서법조동사 사용 빈도 연구에서 나온 결과와 마찬가지로 한국인 화자의 논문에 비해 원어민 화자의 논문에서 훨씬 두드러지게 나타난다는 것을 알 수 있다. 또한 EN에서는 ‘our’도 상위에서 확인되고 있는데 이는 ‘we’가 EN에서 압도적인 중요도를 차지하는 것과 일맥상통한다고 볼 수 있다. 반면에 KN에서는 ‘his’가 중요도 있는 어휘로 나타났고 어느 글에서나 많이 나타나는 정관사 ‘the’이지만 서로의 코퍼스를 참조한 중요 어휘에서는 KN에서 압도적으로 더 많이 나타난 것으로 확인되었다. 그리고 KN에서는 ‘sentence, VP, adjunct NP, structure, predicate, object, ellipsis’와 같이 구문론적인 어휘들이 상당히 많이 나타나고 있고 EN에서는 ‘dialect, variation, variant, dialects’ 등과 같은 사투리와 변이

에 대한 어휘들이 나타났다. 마지막으로 이름으로는 EN에서는 'Dickens, Jones, Anita, Brandy, Burne, Labov'가 나타났고 KN에서는 'John, Whitman, Bradley, Margery, Lacan'이 나타났다.

V. 결론 및 제언

지금까지 본 연구는 코퍼스를 이용하여 원어민 화자와 한국인 화자간의 영어 관련 학문(영문학, 영어학, 영어교육학)의 학술 논문을 가지고 언어적 특성을 다양한 측면에서 비교 분석해 보았다. 이는 세계화 속에서 하루가 다르게 발전하는 학문 분야에 발맞추어 영어로 된 다양한 학술 논문들이 쏟아져 나오고 있는 가운데 EFL 환경에서 외국어로써 영어를 학습하는 우리나라 사람들의 학문적 글쓰기에 있어서 원어민의 글쓰기와 비교하여 보완점을 찾고 어려움을 조금이나마 해소하고자 한 것이다. 분석을 위해서 각각 2008년에서 2012년 사이에 발표된 학술지의 학술 논문을 학문별로 균형을 맞추어 수집하여 코퍼스를 구축하였다. 코퍼스 프로그램은 AntConc3.2.4w를 주로 사용하였다. 그 결과를 요약하면 다음과 같다.

먼저 영어 학습에 있어서 중요한 어휘의 측면에서 살펴보았다. 어휘는 크게 내용어와 기능어로 나뉘는데 각각의 코퍼스로 어휘 목록을 추출하여 상위 빈도로 나타나는 내용어를 중심으로 살펴보고 다시 어휘 목록에서 기능어만을 따로 비교해 보았다. 전반적으로 같은 학문 분야에서 나타난 어휘들이라 보니 공통적으로 확인되는 어휘들이 많았는데 영어 관련 학문과 관계있는 'English, student, language, learning, learners, words, writing, speakers' 등이 많이 사용된 것으로 확인되었다. 차이점이라고 한다면 원어민 코퍼스에서는 동사로는 'use, see, change'가 많은 빈도수로 사용되었고 형용사로는 'social, new, different, particular'등이 확인되었다. 반면에 한국인 코퍼스에서는 동사는 'use'만, 형용사는 'new'와 'different'만 확인되었다. 아무래도 두 코퍼스 모두 상위 빈도에는 명사가 대부분을 차지하지만 원어민 코퍼스의 경우 명사 외에도 동사, 형용사가 골고루 나타나는 반면 한국인 코퍼스는 명사가 대부분을 차지했다. 본 논문에서는 상위 빈도에서 나

타나는 어휘들을 주로 살펴보았지만 좀 더 목록을 확대하여 살펴본다면 두 코퍼스 간 내용어의 사용 양상의 차이를 좀 더 뚜렷하게 알 수 있을 것으로 생각된다.

다음으로 어휘의 측면에 있어서 내용어 못지않게 중요한 기능어의 빈도수를 통한 사용 양상을 비교해 보았다. 내용어 만으로도 어느 정도 내용이 파악은 되지만 문법적인 역할을 담당하며 문장을 매끄럽게 연결해주는 기능어 없이는 자연스럽게 논리적으로 완성된 글을 만들기 어렵기 때문에 결코 간과해서는 안 되는 것이다. 역시 내용어와 마찬가지로 상위 빈도로 나타나는 기능어를 비교해 보았는데 기능어는 내용어와 달리 그 수가 한정되어 있기 때문에 전체적으로 거의 비슷하게 나타난 것으로 확인되었다. 'the, of, and, in, to, a'의 순으로 많이 나타났으며 그 외에도 'for, with, by'와 같은 전치사와 'this, it'과 같은 대명사 등이 나타났다. 가장 널리 사용되는 BNC 코퍼스와 비교했을 때도 비슷한 결과를 보였다. 다만 원어민 코퍼스에서만 나타난 기능어로는 전치사 'from'이 있었고 한국인 코퍼스에서만 나타난 기능어로는 대명사 'his'가 있었다. 또한 두 코퍼스 모두 BNC 코퍼스에서는 상위에서 나타난 인칭 대명사 'you, we, they'등은 나타나지 않았다. 이는 두 코퍼스의 글 종류가 학술 논문이라는 것에 기인하는 것으로 보여 진다.

다음으로 기능어 중에서도 학술 논문의 특성상 정확한 내용의 전달을 위해 무엇인가를 지칭하는 역할을 하는 대명사 또한 중요하기 때문에 따로 비교해 보았다. 대명사에도 여러 종류가 있지만 본 논문에서는 두 코퍼스 간에 대표적으로 많이 나타나는 인칭 대명사와 지시 대명사를 중심으로 살펴보았다. 먼저 인칭 대명사의 경우 'I, they, he'의 경우 두 코퍼스 간에 비슷하게 나타났다. 반면에 큰 차이점으로는 'we'에서 나타났다. 원어민 코퍼스에서 두 배 이상 더 많이 사용한 것으로 확인되었는데 이는 원어민의 학술 논문의 경우 공동연구가 더 많은 것에서 기인한 것이 아닐까 생각하여 두 코퍼스

각각 공동연구가 얼마나 되는지 함께 확인해 보았다. 10개의 개수 차이가 났는데 생각보다 많은 수의 차이가 아닌 것으로 판단해 볼 때 원어민 화자와 한국인 화자 사이에 글을 쓰는 데 있어 사고방식이나 스타일의 차이가 있는 것이 아닐까 생각해 보았다. 그러나 일반화하기에는 다소 무리가 있으므로 좀 더 많은 연구가 이루어질 필요가 있을 것이다. 다음으로 지시 대명사의 경우 'it'과 'this'가 비슷한 비율로 많이 나타났고 차이점이라고 한다면 'these'와 'those'인데 원어민 코퍼스에서 'these'의 사용이 'those'의 사용보다 세 배 이상 나타났다. 한국인 코퍼스의 경우는 두 배 이상 나타나서 두 코퍼스에서 모두 'these'의 사용이 'those'의 사용보다 많은 것은 공통적인 결과이나 그 사용 빈도수면에서는 확실히 차이가 남을 알 수 있었다.

다음으로 어떻게 사용하느냐에 따라서 의미가 상당히 달라질 수 있기 때문에 의사소통에 있어서 중요한 역할을 하는 서법 조동사의 사용 양상을 비교해 보았다. 그 결과 전체적으로 원어민 코퍼스에서 서법 조동사의 사용 비율이 한국인 코퍼스에 비해 조금 더 많이 나타났다. 또한 두 코퍼스 모두 'can'의 사용이 가장 많았고 'can, may, will'과 같은 현재형이 'could, might, would'와 같은 과거형보다 더 많이 사용되었다. 반면에 뚜렷한 차이라고 한다면 원어민 화자와 한국인 화자 간에 'could'사용에 있어서 많은 차이로 원어민 화자가 더 많이 사용하고 있었고 이는 'might'과 'would'의 경우도 마찬가지로 결과를 보였다. 이 결과를 토대로 확실히 단정 짓기는 무리가 있겠지만 서법 조동사의 과거형 사용에 있어서는 분명한 차이를 보임은 확실하므로 한국인 화자가 서법조동사의 사용에 있어서 현재형을 좀 더 선호하고 과거형의 사용에서 다소 어려움을 느낀다고 생각되어 진다. 이어서 의무와 확신을 나타내는 'should'와 'must'를 보면 원어민 화자의 경우 둘의 사용 비율이 비슷한 반면 한국인 화자의 경우 'should'의 사용이 두 배 이상 많았다. 이로써 한국인 화자가 강한 의미의 'must'보다는 보다 보편적으로

사용할 수 있는 'should'를 더 선호하는 것으로 생각해 볼 수 있다. 이러한 결과를 바탕으로 의사소통에 있어 굉장히 중요한 역할을 하는 서법 조동사의 교육을 하는데 있어서 어느 한쪽만을 더 선호해서 치우친 경향을 보이지 않도록 다양한 쓰임을 접할 수 있도록 하고 그 사용을 권장하여 좀 더 정확한 의미전달을 할 수 있도록 교육할 필요가 있을 것이다.

다음으로 어휘 묶음(lexical bundle)의 사용 양상을 비교해 보았다. 예전에 비해 요즘 들어 원어민과 같은 자연스러운 영어를 구사하기 위해서 관용어나 연어, 어휘 묶음 등의 단어 연쇄체를 익히고 사용하는 것이 중요하게 대두되고 있는데 코퍼스 언어학이 발달하고 컴퓨터가 발달하면서 다양한 프로그램이 개발되어 활발히 그 연구들이 이루어지고 있다. 그에 발맞추어 본 논문에서도 원어민 화자와 한국인 화자의 학술 논문에서 나타난 다양한 어휘 묶음들에 대한 비교 연구를 하였다. 이를 통해 조금이나마 어휘 묶음의 사용을 통한 자연스럽게 매끄러운 학문적 글쓰기를 할 수 있도록 도움이 되고자 하였다. 어휘 묶음에 대한 연구를 크게 세 부분으로 나누어 볼 수 있는데 첫째로 많은 빈도수로 나타나는 어휘 묶음을 확인하고 각각을 학문 분야에 따라 빈도수와 나타난 텍스트 수를 확인하였다. 둘째로는 Cortes의 분류 기준에 따라 구조적인 측면에서 분류해 비교하였고 마지막으로 역시 Cortes의 분류 기준에 따라 기능적인 측면에서 분류해 비교하였다.

먼저 사용 빈도의 비교 분석 결과를 보면 두 코퍼스에서 나타난 어휘 묶음의 유형과 어휘 수, 전체 빈도수의 합과 텍스트의 합에서는 큰 차이를 보이지 않았다. 반면에 확인한 수치에는 차이가 있었는데 원어민 코퍼스에서 영문학과 영어교육학에 비해 영어학에서 훨씬 많은 어휘 묶음을 확인할 수 있었고 한국인 코퍼스에서는 영문학과 영어학에 비해 영어교육학에서 더 많은 어휘 묶음을 확인할 수 있었다. 나타난 어휘 묶음 가운데 상위의 빈도로 나타난 것들을 비교하였는데 비슷하게 나타난 것들이 많이 있었다. 우선 원어민

민 코퍼스에서는 ‘at the same time’이 가장 많이 나타났고 한국인 코퍼스에서는 ‘on the other hand’가 압도적인 수치로 많이 나타났다. 또 두 코퍼스에서 모두 많이 나타나고 있는 어휘 묶음으로는 ‘the end of the, in the case of, as well as the, as a result of, at the end of, the fact that the, in terms of the, on the basis of, in the form of’등이 확인되었다, 반면에 ‘the ways in which, in the context of, it is important to, at the time of, on the one hand, it is clear that, the extent to which, the use of the’는 원어민 코퍼스에서만, ‘in the process of, the results of the, one of the most, as shown in figure, the beginning of the, is one of the, at the beginning of, in the sense that, the purpose of this’는 한국인 코퍼스에서만 나타났다. 주목할 점은 한국인 코퍼스에서 ‘on the other hand’가 압도적으로 사용된 데 비해 ‘on the one hand’는 상위에 나타나지 않은 반면 원어민 코퍼스에서는 상위에 나타나고 있었다. 다음으로 각 분야 당 많이 나타난 어휘 묶음을 확인해 보았는데 전체적으로 영문학 분야에서 확인되는 어휘 묶음이 적은 만큼 많은 종류가 확인되지는 않았다. 원어민 코퍼스에서는 ‘as a means of, in the face of, in the first place’등이 확인되었고 한국인 코퍼스에서는 ‘can be seen as, in the form of, it is true that’등이 확인되었다. 영어학 분야에서는 종류의 측면에 있어서는 원어민 코퍼스에서 현저히 많은 것으로 확인되었다. 이 결과로 단정 짓긴 어렵지만 주제의 특성상 정확한 데이터를 전달하기 위해 다른 학문과 달리 미사여구 보다는 고정된 어휘 묶음을 사용한 간결한 문장을 선호하는 것으로 볼 수도 있을 것이다. 나타난 어휘 묶음으로는 원어민 코퍼스에서는 ‘a great deal of, a wide range of, are more likely to, can be used to, it is difficult to, it is possible that, on the part of, to account for the, to the extent that’등의 다양한 어휘 묶음이 나타났고 한국인 코퍼스

에서는 ‘in addition to the, in spite of the, in the same way, in the sense that, the subject of the, this paper is to’등이 나타났다. 마지막으로 영어교육학 분야에서는 영어학에서와는 달리 한국어 코퍼스에서 더 많은 종류의 어휘 묶음이 나타났는데 원어민 코퍼스에서는 ‘at the beginning of, in a way that, in the process of, in the role of, the extent to which, what it means to’등이 나타났고 한국인 코퍼스에서는 ‘as one of the, at the beginning of, can be said that, for the purpose of, the results of the, the use of the, this study is to, with regard to the’등이 나타났다.

다음으로 어휘 묶음의 구조에 따른 사용 양상을 비교한 결과 전체적으로 두 코퍼스 모두 ‘the beginning of the, one of the most’와 같이 ‘in terms of the, in the case of’와 같이 전치사 ‘of’가 포함되어 있는 명사구와 ‘of’가 포함되어 있는 전치사구의 구조가 많이 나타나는 것으로 확인되었다. 이 점은 Cortes(2004)의 결과와도 같았다. 또한 ‘despite the fact that, in addition to the’와 같이 다른 전치사구 역시 많이 확인되었다. 그 외에도 ‘is one of the’와 같이 동사 다음에 명사구인 보어로 이루어진 구조, ‘are more likely to, can be used to’와 같이 서법 조동사나 be동사와 뒤 이어 나오는 명사구나 형용사구의 보어로 이루어진 구조, ‘it is necessary to, it is possible to(that)’과 같이 it을 주어로 뒤이어 be동사와 형용사, 그리고 절로 이어지는 구조가 확인되었다. ‘an important role in, difference between the two, the relationship between the’과 같이 ‘of’가 포함되지 않은 명사구의 구조와 ‘the results showed that, this paper is to, this study is to’와 같이 명사 혹은 대명사를 주어로 뒤이어 동사와 보어가 있는 구조의 경우 한국인 코퍼스에서만 확인되었다.

마지막으로 어휘 묶음의 기능에 따른 사용 양상을 비교하였는데 전체적으로 큰 범주에서 지시 어휘 묶음이 가장 많이 나타났고 뒤이어 텍스트 구성

어휘 묶음과 견지 어휘 묶음 순으로 많이 나타났다. 세부적으로 보면 시간 어휘 묶음과 양화 어휘 묶음은 원어민 코퍼스에서 좀 더 많이 나타났고 장소 어휘 묶음과 묘사 어휘 묶음은 두 코퍼스에서 비슷하게 나타났다. 다음으로 텍스트 구성 어휘 묶음에서는 틀형성의 기능을 하는 어휘 묶음이 가장 많이 나타났다. 견지 어휘 묶음의 경우 두 코퍼스 모두 비슷한 수로 나타났다.

이렇듯 두 코퍼스 간의 어휘 묶음 결과를 종합적으로 보자면 비슷한 측면도 많이 가지고 있는 반면에 분명한 차이들도 많이 있었다. 다른 연구들의 결과에서와 마찬가지로 영어 교육이 발달함에 있어 원어민에 비슷한 실력을 가진 비원어민 영어 학습자들이 많아짐에 따라 두루 사이의 비교에 있어서 너무나 상이한 결과를 보이지는 않는 것이 사실이다. 그러나 사용하고 있는 종류에 있어서 원어민의 경우 좀 더 다양한 것을 확인할 수 있었고 한국인 화자가 선호하는 어휘 묶음의 사용에 있어서 더 편중되는 측면도 볼 수 있었다. 이런 만큼 영어 교육에 있어서 좀 더 다양한 어휘 묶음을 다룸으로써 다양성을 키워줄 필요가 있을 것이다.

다음으로 본 논문에서는 결속력 있는 글을 쓰기 위하여 반드시 필요한 요소인 접속(부)사에 대한 사용 양상을 비교 분석해 보았다. 중요한 요소인 만큼 다양한 연구들이 많이 이루어져 왔는데 본 논문의 결과를 바탕으로 다른 연구 결과들과 비교해 보았다. 그 사용 빈도와 문장 안에서 나타나는 위치를 중점적으로 살펴보았다. 먼저 역접의 의미를 갖고 있고 전반적으로 여러 글에서 많은 빈도수를 보이는 'however'의 경우 빈도수의 경우 두 코퍼스 간에 큰 차이를 보이지 않지만 사용 위치에는 큰 차이를 보였다. 원어민 코퍼스의 경우 문두와 문중의 사용 비율이 비슷한 반면 한국인 코퍼스의 경우 문두에서의 사용이 문중 보다 두 배 이상 많았다. 고헤(2008)의 경우 'however'의 사용이 원어민 코퍼스의 경우 문중이 압도적으로 많았고 한국

인 코퍼스의 경우 문두가 압도적으로 많았다고 제시하였다. 본 논문의 결과와 비교했을 때 원어민 코퍼스의 경우에는 다소 차이가 있었지만 한국인 코퍼스의 경우 비슷한 결과를 나타냈다. 영어를 전공하는 한국인 화자의 학술 논문에서도 이러한 결과가 나온 것으로 미루어 영어 교과서에서 ‘however’가 문두에 훨씬 더 많이 나타난 것이 우리가 영어 교육을 받아오기를 ‘however’를 문두에 쓰는 경우를 더 많이 볼 수밖에 없었기 때문에 자연스럽게 한국어 화자들이 문두에 사용하는 것을 더 선호하는 경향이 있는 것으로 생각해 볼 수 있다. 영어 교육에 있어 한쪽으로 편중된 위치만을 학습시키기보다 다양한 가능성을 함께 제시할 필요가 있을 것이다. 다음으로 ‘so’의 경우 두 코퍼스에서 비슷한 빈도로 나타났고 사용 위치는 문중에서 압도적으로 나타났다. 다른 연구 결과와 비교해보면 구선화(2012)는 한국인 화자가 원어민 화자에 비해 ‘so’를 압도적으로 많이 사용하고 있다고 하였는데 이는 본 논문의 결과와 다소 차이가 있었다. 이러한 경향을 좀 더 확실히 확인하기 위해서는 좀 더 다양한 연구가 필요할 것으로 보인다. ‘therefore’과 ‘thus’의 경우 두 코퍼스 간의 빈도수는 크게 차이하지 않다. 그러나 구선화(2012)는 ‘thus’의 경우는 원어민이 더 많이 사용한다고 제시하였다. 이 역시도 결과가 본 논문과 상이하여 추가 연구가 필요할 것이다. 다음으로 ‘therefore’의 위치의 경우 원어민 코퍼스의 경우 문중이 압도적으로 많았고 한국인 코퍼스의 경우 문두가 압도적으로 많았다. 이는 고광윤(2008)의 연구 결과와 같게 확인되었다. 이렇듯 다양한 종류의 글에서 접속부사의 사용 위치에 대한 차이가 남으로써 이를 해소하기 위해서는 영어 교육에 있어서 편중된 위치 사용이 아닌 다양한 사용의 용례를 보임으로써 학습자가 자연스럽게 익힐 수 있도록 할 필요가 있고 원어민이 작성한 좀 더 다양한 글을 접할 수 있는 기회를 많이 마련해줘야 할 것이다.

마지막으로 참조 코퍼스를 통한 중요 어휘를 비교 분석해 보았다. 참조 코

퍼스로는 BNC 코퍼스와 서로의 코퍼스를 이용하였다. 먼저 공통적으로 내용어로는 'student, teacher, learner'과 같은 교육 측면의 어휘가 나타났고 기능어로는 'as, of, in'이 나타났다. 또한 원어민 코퍼스에서는 'dialect, variation, texts, frequency'등의 어휘가 나타난 것으로 최근 연구 경향이 사투리나 변이, 코퍼스 언어학이 많이 이루어지고 있다고 생각할 수 있고 한국인 코퍼스의 경우 'feedback, proficiency, native'와 같은 어휘들로 이루어져 원어민에 가까운 언어 구사를 목표로 하는 영어 교육이 많이 강조되고 있음을 생각해 볼 수 있고 'sentence, syntactic, constructions, clauses, adjunct, VP, NP'등과 같이 구문론적 어휘들도 많이 나타났다. 또한 서로의 코퍼스를 참조하여 나타난 중요 어휘를 비교해 본 결과 두드러진 차이로는 본 논문의 연구에서도 확인했듯이 원어민 코퍼스에서 한국인 코퍼스에 비해 두드러진 중요 어휘로 'we, these, our, would'등이 확인되었고, 한국인 코퍼스에서는 'the, Korean, feedback, EFL, proficiency, his'등이 확인되었다.

이와 같이 본 논문에서는 다양한 측면에서 두 코퍼스 간의 차이를 비교 분석해 보았다. 본 연구 결과를 토대로 영어 교육적 측면에서 몇 가지 시사점을 제공해 줄 수 있다. 영어를 외국어로써 배우는 우리나라의 환경에서 언어에 대한 노출이 원어민에 비해 훨씬 적은 만큼 원어민처럼 영어를 자유자재로 구사하기에는 많은 한계점이 따르는 것이 사실이다. 이러한 차이를 극복하기 위해서 영어 교육에 있어서 다양한 매체를 통한 다양한 장르의 진정성 있는 영어를 접할 수 있는 기회를 최대한 많이 제공해 주어야 할 것이다. 이를 통해 영어 학습자가 기존의 배워온 정해져 있는 공식에 따른 인공적인 영어가 아닌 자연스러운 영어 사용에 익숙해 질 수 있을 것이다. 또한 영어를 가르치는 교수자의 입장에서 먼저 본 논문의 결과에 나타난 것과 같은 원어민과 비원어민의 글에 나타나는 차이점들을 인식하고 제대로 가르친다

면 영어를 배우는 학습자의 입장에서 결코 쓰기 쉽지 않은 학문적인 글을 쓰는데 있어서 혹은 더 나아가 다른 장르의 글을 쓰는 데 있어서도 훨씬 많은 도움을 받을 수 있을 것이라고 생각한다.

참고 문헌

- 김경아. 2007. 원어민과 비원어민의 논문 초록에 나타난 축약 관계사절 연구. 석사 학위 논문. 계명 대학교
- 고광윤. 2005. 한국적 영어학과 영어 코퍼스의 활용. *English language and Linguistics* 19. 1-19
- 고광윤. 2008. 교과서 영어의 영어다움에 대한 코퍼스 기반 분석 : 기능어의 쓰임을 중심으로. *English language and Linguistics* 26. 1-19
- 고광윤. 2009. 코퍼스 분석을 위한 프로그램의 선택 Wordsmith Tool 4.0과 MonoConc Pro 2.2를 중심으로. *English language and Linguistics* 27. 1-22
- 고수원. 2010. 한국인 학자와 영어 원어민 학자의 논문 영문 초록 비교 분석. *영어어문교육* 16권 1호 189-208
- 구선화. 2012. 영어 원어민과 한국인 영어 학습자의 학술지 논문 코퍼스에 나타난 결과 접속 부사 사용 비교. 석사 학위 논문. 이화 여자 대학교
- 김경숙. 2000. 한국과 미국의 영문 초록 텍스트의 담화 분석. 석사 학위 논문. 숭실 대학교
- 김준기. 2007. 세계 영어의 지역적 차이에 대한 말뭉치 분석 연구. *새한영어영문학* 제49권 2호 143-155
- 김준기. 2011. 현대 미국영어 코퍼스에 나타난 서법 조동사와 부사 연어 분석 연구. *새한영어영문학* 제53권 1호 175-200
- 서지연. 2007. 영어 교육 관련 학술 논문 영문 초록 분석 : 수사 구조와 담화 표시어를 중심으로. 석사 학위 논문. 한양 대학교.
- 성일호. 2010. 연어의 단어 결합력 측정 코퍼스 수업안 개발. *신영어영문학* 47집 225-243

- 성일호. 2009. 코퍼스 분석을 통한 조동사 활용 수업안 개발. *신영어영문학* 43집 235-257
- 이은주. 2008. 영어 교육과 응용 언어학 분야에서 수행된 코퍼스 기반 연구의 분석. *영어교육* 63권 2호 283-306
- 이재희. 2004. 학술지 논문 영문 초록의 담화 구조와 언어 특징 : EAP 교육을 위한 탐색. *Foreign Languages Education* 11(1) 89-109
- Allen, David. 2009. Lexical bundles in learner writing: An analysis of formulaic language in the ALESS Learner Corpus. *Komaba Journal of English Education* 1.1: 105-127.
- Biber, Douglas and Federica Barbieri. 2007. Lexical bundles in university spoken and written registers. *English for Specific Purposes* 26: 263-286
- Byrd, Pat and Averil Coxhead. 2010 *on the other hand*: Lexical bundles in academic writing and in the teaching of EAP. *University of Sydney Papers in TESOL*. 31-64
- Chen, Qi and Guang-chun Ge. 2007. A corpus-based lexical study on frequency and distribution of Coxhead's AWL word families in medical research articles (RAs). *English for Specific Purposes* 26: 502-514.
- Chen, Yu-Hua and Paul Baker. 2010 Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology* 14.2: 30-49.
- Cortes, Vivian 2004. Lexical bundles in published and student disciplinary writing : Examples from history and biology. *English for Specific Purposes* 23: 397-423.
- Jhang, Se-Eun, and Shin Chul Hong. 2012. A corpus-based Analysis of

Journal Article Abstracts Written in English by Korean Authors : A comparison of English Literature and English Linguistics. A paper presented at The 1st World Congress of Scholars of English Linguistics, Seoul, Korea.

Salazar, Danica Joy Lorenzo. 2011. *Lexical bundles in scientific English : A corpus-based study of native and non native writing*. Doctoral dissertation. University of Barcelona.

Abstract

A corpus-based analysis of linguistic characteristics in academic articles written by English-native speakers and Koreans

Song, Kyung-jin
Majoring in English Education
Graduate School of Education
Sungshin Women's University

It would be far from an easy goal for Korean learners to become able to speak or write like native speakers in an EFL environment. As educational methods have developed and diversified, more and more academic articles have been published in this field. It is important to express one's thoughts and the results of one's research logically. Therefore, in order to help Korean learners write academic articles more efficiently and naturally, in this study I have compared, based on corpus linguistics, various linguistic characteristics in academic articles written by English-native speakers and Korean authors. The purpose of this study was to examine and identify the differences and similarities between English-native speakers and Koreans in terms of academic

writing, and to discuss how to apply these findings to English education. For this study, among many other disciplines, academic articles on English literature, English linguistics and English education were collected and analyzed. The results of this study are as follows.

First, in an analysis of vocabulary in regard to content words, nouns appeared similarly in the two corpora, but native English speaking authors used a more diverse range of verbs and adjectives than Korean authors did. Then, in regard to function words, compared with the British National Corpus(BNC), these same results showed that articles and prepositions usually appeared at the same rate in the two corpora.

Second, among function words, pronouns were separately analyzed. The personal pronoun, 'we' in academic articles by native English speakers appeared twice as often as in articles written in English by Koreans. And in terms of the demonstrative pronouns 'these' and 'those', native English-speaking authors used 'these' more than Korean authors did.

Third, modal verbs, which play a significant role in communication, were examined. On the whole, modal verbs were used a little more often in academic articles of English-native authors, and, specifically, English-native authors used the past form of modal verbs such as 'could', 'would', and 'might' much more than Korean authors did. Finally, 'should' and 'must', which represent obligation, were used even more often by Korean authors than by native English speaking authors.

Fourth, lexical bundles such as 'in terms of the' and 'on the other hand' were examined. These were analyzed based on their frequency, structure, and function. Concerning frequency, in the case of academic articles by native English speaking authors, lexical bundles appeared in papers on English linguistics more than in other fields, and in the case of journal articles by Korean authors, lexical bundles appeared in English education articles more than in other fields. 'At the same time' appeared most frequently in academic articles of native English speaking authors and 'on the other hand' appeared most frequently in the journal articles of Korean authors. Regarding structure, noun and prepositional phrases which include the preposition 'of', appeared most often, and prepositional phrases including other prepositions also appeared frequently. And last, as for function, by and large, referential bundles were frequently used in the two corpora. In terms of sub-category, the native English speaking authors used quantifying bundles a lot more often than Korean authors did. Framing, a sub-category that functions as a text organizer, was the most frequently used function in the two corpora.

Fifth, conjunctions and conjunctive adverbials needed for writing cohesion were investigated. Regarding 'however', there were not many differences between the two corpora, but in terms of position, the native English speaking authors used 'however' similarly in both initial and medial positions. On the other hand, Korean authors used 'however' far more often in a initial position than in a medial

position. In the case of 'therefore' and 'thus', there were also not many differences regarding frequency. But in terms of position, the great majority of native English speaking authors used 'therefore' in a medial position, while Korean authors used 'therefore' in an initial position.

Sixth, key words in the two corpora were examined, using the BNC as a reference. The results showed that words such as 'student', 'teacher', 'learner', 'as', 'of', 'in' were used commonly by both groups. On the other hand, in academic articles by native English speaking authors, key words included 'dialect', 'variation', 'texts', and 'frequency', while for Korean authors, 'feedback', 'proficiency', 'native', 'syntactic', 'constructions', 'clauses', and 'adjunct' appeared often. The results of key words present in each corpus also showed that 'we', 'these', 'our', and 'would' appeared frequently in academic articles of native English speaking authors and that 'the', 'Korean', 'feedback', 'EFL', 'proficiency', and 'his' were common in those of Korean authors.

Based on these findings, this study suggests that more opportunities to learn authentic English through diverse genres need to be given to English learners in order to overcome the differences between native English speakers and Koreans in using English. And also, if educators recognize these differences and teach their students properly, these findings will be helpful for training English learners to write academic articles as well as various other forms of writing.