

원자력과학공학 학술 논문에 나타난 기능적 어휘다발 분석

Functional Lexical Bundles in Nuclear Science and Engineering Research Articles

남대현
사이버한국외국어대학교 영어학부

Daehyeon Nam(dnam@cufs.ac.kr)

요약

본 연구의 목적은 영어로 작성된 원자력과학공학 학술 논문에 나타나는 어휘다발을 담화기능에 따라 분류한 후, 분류된 어휘다발이 일반 학술 논문에 나타나는 어휘뭉치와 비교하여 어떤 특징을 나타내는지 분석하는데 있다. 이를 위해 원자력과학공학 논문의 텍스트를 수집하여 제작한 약 1백만 단어의 코퍼스에서 기능적 어휘다발을 추출한 후 이를 75만 단어 크기의 일반 학술 논문 코퍼스에 나타난 어휘다발 분포와 빈도를 카이제곱 검정과 표준화 잔차를 사용하여 비교하였다. 그 결과 원자력과학공학 분야에서는 일반 학술 논문과 비교했을 때 저자태도와 관련한 어휘다발이 주로 사용되었고, 어휘다발 사용에 있어서는 다양성이 결여된 어휘다발 사용이 나타나 동일한 타입의 어휘다발을 '재사용'하는 모습을 보여주었다. 이러한 연구결과를 바탕으로 원자력과학공학 학술목적영어 교육에 대한 교육적 함의와 후속연구의 방향에 관하여 제언하였다.

■ 중심어 : | 코퍼스 | 어휘다발 | 원자력과학공학 | 학술목적영어 | 과학기술영어 |

Abstract

This study aims to functionally classify lexical bundles appearing in academic papers on nuclear science and engineering written in English and then analyze the lexical bundles' characteristics compared to those appearing in general academic papers. To this end, the texts of nuclear science and engineering papers were collected and produced as a corpus(c. 1 mil. tokens). Then they were statistically compared through Chi-square tests and standardized residuals with the corpus of general academic papers(c. 750,000 tokens). The results revealed that, compared to general academic papers, the bundles in the stance lexical bundle category were mainly used among the functional lexical bundle in nuclear science and engineering. The use of the lexical bundles lacked much variety. The same type of lexical bundles was 're-used' and 'recycled'. Based on these research results, educational implications for English for Academic Purposes and the further direction of follow-up research were discussed and suggested.

■ keyword : | Corpus | Lexical Bundle | Nuclear Science and Engineering | English for Academic Purposes | English for Science and Technology |

* 이 연구는 2021학년도 사이버한국외국어대학교 교내학술연구비의 지원에 의하여 이루어진 것임.

접수일자 : 2021년 10월 01일

심사완료일 : 2021년 10월 20일

수정일자 : 2021년 10월 19일

교신저자 : 남대현, e-mail : dnam@cufs.ac.kr

I. 서론

원자력과학공학이라는 학문은 과학기술 관련 학문이 그러하듯이 국제적인 학술 교류와 협력이 필요한 분야이다. 영어는 이러한 국제적이고 전문적인 커뮤니케이션의 중추적인 역할을 담당해 왔는데, 실제로 영어는 지난 20여 년 동안 전 세계적으로 학술 연구를 위한 엄청난 사용량 증가로[1], 이제는 학계의 '링구아 프랑카'(lingua franca)가 되었다[2]. 과학기술의 발전 속도가 점점 빨라짐에 따라 과학 및 공학 분야에 종사하는 연구자들의 가장 큰 관심은 대량으로 생산되는 새로운 연구결과를 해당 분야의 학계에 가장 신속한 방법으로 발표하는데 있다. 이를 위한 효과적인 수단으로 연구자들은 국제 학술대회 발표나 학술 논문을 출판하여 연구의 결과를 발표하는데, 이때 연구자의 능숙한 영어의 사용은 연구 결과 확산에 필수 불가결한 요소로 자리매김하였다. 특히, 많은 과학기술 논문들이 영어를 기반으로 한 학술지를 통해 발표되는 이유는 과학적 발견이 더 넓은 독자층에 전파되고 공유되어야 하는 목적 때문이다. 이러한 이유로 영어교육 연구자나 교육자들은 학술목적영어(English for Academic Purposes, ESP) [3-11]와 특수목적영어(English for Specific Purposes, ESP)에 관한 연구를 활발히 진행해 왔고[11-18], 국내에서도 관광, 화학공학, 항공안전, 컴퓨터과학과 관련한 연구가 수행되었다[19-22].

영어교육 분야에서 학술목적영어와 특수목적영어에 대한 연구와 교육이 활발해지고, 다량의 데이터를 단시간에 효율적으로 처리할 수 있는 컴퓨터 기술이 발달함에 따라 이 분야에 대한 코퍼스언어학(Corpus Linguistics)을 활용한 연구가 수행되어왔다. 이러한 언어학 연구방법의 발전과 더불어 코퍼스 기반 영어교육에서 두 개 이상의 단어가 서로 결합하여 사용되는 표현 또는 어휘다발에 대한 관심과 연구가 활발히 진행되고 있다. 그 이유 중 하나는 학습자가 외국어를 사용할 때 어휘다발은 언어처리의 자동화를 증대시켜 보다 유창한 언어 사용을 가능하게 한다는 장점이 있을 뿐 아니라 학술목적영어와 특수목적영어의 중요성은 해당 분야에 대한 학문적 관심뿐 아니라 이를 응용하는 관련 산업 발전에 따라 크게 좌우되기 때문이다[3]. 실제로 우리나라에서

원자력과학공학은 과학기술 연구의 중요한 한 축을 담당하고 있고, 이러한 학문적 중요성을 고려할 때 원자력과학공학에 사용되는 학술영어에 관한 교육적 함의 도출을 위한 기초연구가 필요한 실정이다. 이에 본 연구의 목표는 다음과 같다:

- 원자력과학공학에 사용되는 어휘다발 (lexical bundle)의 특징을 그 담화 기능별로 분석
- 분석 결과를 통한 원자력과학공학 학술영어 교육에 사용할 수 있는 교육적 함의를 제안

II. 이론적 배경

코퍼스 언어학에서 어휘다발 연구의 목적은 특정한 사용역이나 장르에 '반복적으로 나타나는 연속된 단어 연결체'(recurrent consecutive word sequence)로 기술될 수 있는 언어학적 단위를 찾아내고 그 구조적 기능적 부류와 분포를 연구하는 것이다(예 [23][24]). 일반적으로 어휘다발은 코퍼스 내에서 1백만 단어당 나타나는 횟수의 비율을 기준으로 컴퓨터 프로그램을 통해 의미상 맥락을 고려하지 않고 기계적으로 추출된다. 이러한 이유로 많은 경우 어휘다발은 문법 구조상 불완전한 형태를 취할 수밖에 없다. 그럼에도 불구하고 어휘다발은 텍스트 상에서 담화 프레임 연결하고 구성하는 '기능적(functional)' 역할을 담당한다[25]. 예를 들어, 텍스트의 담화를 구성하는 기능적 어휘다발은 지시적(referential), 저자태도(stance), 그리고 담화조직(discourse organizing)이라는 세 가지 범주로 구분될 수 있다. 지시적 기능을 보이는 어휘다발은 물리적 또는 추상적인 개체를 직접 지시하는 어휘다발이거나, 특정한 개체나 개체의 속성을 식별하기 위해 텍스트 내에 포함된 문맥을 지칭하는 역할을 담당한다. 저자태도 어휘다발은 특정 개체에 대한 저자의 태도 및 인식론적 견해를 전달하는 프레임의 역할을 담당한다. 그리고 담화조직과 관련한 어휘다발은 텍스트 내에 존재하는 여러 가지 담화 진행 단계에 있어서 이전 단계에 제시된 담화와 이에 연결되어 뒤따라오는 새로운 담화의 관계를 엮는 역할을 담당한다(어휘다발 분류에 대한 자세한 예는 III. 방법론 참조).

기능적 어휘다발에 연구에서는 어휘다발이 특정한 장르 또는 사용역에 따라 서로 구별되는 담화 기능이 사용되는 특징이 확인되었다. 예를 들어, 구어와 문어에 사용되는 어휘다발을 전면적으로 조사한 Biber, Johansson, Leech, Conrad와 Finegan(1999)[26]는 일반 대화체로 대표되는 구어체 보다는 학술문에 지시적 기능을 담당하는 어휘다발이 더 많이 사용된다고 밝혔다. 또한, 이들에 연구 결과에 의하면 대화체에서는 저자태도와 관련된 어휘다발이 가장 많이 사용되었고, 그 다음으로는 담화조직과 관련된 어휘다발이 많이 사용되었다. 그리고 지시적 기능의 어휘다발이 가장 적게 사용되었다.

Hyland(2008)[27]는 학술지에 출판된 전기공학, 생물학, 경제학, 그리고 응용 언어학 논문에서 사용된 어휘다발의 분포와 그 특징을 연구했다. 이 연구 결과에 따르면, 응용 언어학과 경제학 연구논문에서는 담화조직 기능과 관련한 어휘다발이 전기공학이나 생물학에서 보다 더 많이 사용된다. 또 다른 연구인 Cortes(2004)[28]는 역사 학술문 텍스트의 지시적 어휘다발에는 시간에 관련한 어휘다발 타입은 나타나지 않은 사실을 발견한 반면, 생물학 분야 학술문 텍스트에서는 생물학적 현상인 진화나 발달과정의 여러 단계를 지칭하는 시간 관련 지시적 어휘다발이 다수 사용된 사실을 밝혔다(예, *at the beginning of, at the same time, the end of the, in the life cycle*). 이러한 결과는 동일한 학술문이라는 장르라 하더라도 기능적 어휘다발은 학문 분야의 주제와 방법론에 따라 서로 구별되는 어휘다발이 사용된다는 것을 의미한다. 또 저자태도 어휘다발 중 인식론적 기능을 담당하는 어휘다발은 저자가 보이는 확신의 정도를 조정하는 기능을 담당한다. 생물학 학술 논문 텍스트 분석에 의하면 이 인식론적 기능을 하는 어휘다발 타입이 다수 나타났다(예, *may be due to, it is possible that, it is likely that, is likely to be, are likely to be, are more likely to be, the probability that a*). 그러나 역사 분야 연구 논문 텍스트에서는 이러한 역할을 하는 어휘다발은 나타나지 않아 생물학 연구자들은 4개의 어휘가 연속된 어휘다발보다는 짧은 어휘다발이나, 아니면 완전히 다른 어휘를 사용하여 그 역할을 대치하는 것으

로 여겨졌다. 지금까지 살펴본 선행연구 결과는 기능적 어휘다발의 사용과 분포는 구어와 문어라는 사용역에 따라, 그리고 동일한 사용역(문어)이고 장르(학술문) 내에서도 그 학술 분야에 따라 서로 다른 빈도와 분포로 사용한다는 점을 시사한다.

III. 방법론

1. 분석 대상 코퍼스

이 연구를 위해 두 개의 코퍼스가 사용되었다. 우선 연구 코퍼스(study corpus)인 원자력과공학학 코퍼스(이하 NSE)는 저명 학술지에 출판된 논문을 체계적으로 수집하여 코퍼스를 제작하였다. 코퍼스 제작을 위한 목표 학술지를 선정하기 위해 원자력과공학 원자력공학을 전공으로 하는 대학교수 3명을 각각 인터뷰 하였다. 그 결과에 따라 관련 분야에 관한 양질의 연구를 출판하는 학계의 인지도도 확보할 수 있고 논문에 사용된 영어의 질도 해당 분야의 특성을 충분히 반영할 수 있도록 Journal Citation Index의 상위 10% 이내의 학술지이면서 폭넓은 독자층과 다양한 세부 전공에 대한 학술 논문을 출판하는 학술지인 *Annals of Nuclear Energy, Nuclear Engineering and Technology*, 그리고 *Nuclear Science and Engineering*을 목표 학술지로 선정하였다. 이렇게 선정된 학술지 중 2011, 2012, 2013년에 출판된 학술 논문을 PDF 파일로 내려받아, 이를 텍스트 파일(.txt)로 저장하였다. 이때 논문의 제목, 저자 약력, 초록, 키워드, 각주, 공식, 표/그림과 캡션 또는 참고문헌과 같이 본 연구와 직접적인 관련이 없는 사항은 텍스트 파일 제작 시 포함하지 않았다. 저장된 논문 텍스트 파일은 각각 출판된 권, 호수 별로 하위 코퍼스 폴더에 저장하여 최종적으로 NSE를 완성하였다.

NSE의 비교 분석을 위한 참조 코퍼스(reference corpus)로는 BNC-Baby의 일부가 사용되었다. BNC-Baby는 1억단어로 구성된 British National Corpus에서 추출한 약 400만 단어로 만들어진 균형 코퍼스(balanced corpus)이다. 이 소형 코퍼스는 구어, 신문기사, 소설, 그리고 학술지(논문) 등 총 4개의

하위 코퍼스로 구성되어있는데, 이 하위 코퍼스 중 학술 관련 텍스트를 포함하고 있는 하위코퍼스를 본 연구의 참조 코퍼스로 정하였다(이하 BNC-BaAc).

표 1. NSE와 BNC-BaAc 코퍼스 기본정보

개념	NSE	BNC-BaAc
내용	원자력과공학학 학술논문	일반 학술논문
텍스트 개수	222	30
텍스트 평균 길이(토큰)	4,587.73	25,013.47
타입 개수	16,974	31,337
토큰 개수	1,018,032	750,404

[표 1]에서 알 수 있듯이 NSE는 연구의 효율성과 확장성을 위해 약 1백만 단어의 코퍼스로 제작되었다. 이 크기는 컴퓨터 판독이 가능한 현대적(그리고 최초의) 전자 코퍼스인 Brown Corpus[28]의 크기와 비슷한 크기로서 이후 제작된 소형 코퍼스들은 비교의 편리성을 도모하기 위해 많은 경우 1백만 단어 크기로 제작되기 때문이다.

2. 어휘다발 선정 기준

코퍼스 추출을 위해 고려해야 할 사항은 어휘다발의 길이와 출현 빈도수인데, 우선 어휘다발의 길이는 4개의 단어가 연속된 어휘다발로 정하였다. 그 이유로는 이론적으로 어휘다발의 길이는 단어가 2개 또는 그 이상의 단어로 연결된 길이가 가능한데, 실제로 4개의 단어로 이루어진 어휘다발은 3개의 단어로 이루어진 어휘다발의 일부를 포함할 수 있고 4개의 단어로 이루어진 어휘다발은 5개의 단어로 이루어진 어휘다발보다 더 자주 나타나기 때문이다[24][27]. 또한 연구의 효율성이라는 관점에서 볼 때 4개의 단어로 이루어진 어휘다발을 추출할 때 얻어지는 어휘다발의 개수가 연구자가 수동으로 확인하고 분류하여 처리 가능한 적당한 수로 얻어지기 때문이다[30][31].

일반적으로 코퍼스 분석에 있어서 단어나 또는 어휘다발의 출현 빈도수는 코퍼스의 크기에 따라 민감하게 달라진다. 따라서 크기가 다른 두 개의 코퍼스를 비교할 경우 그 결과가 왜곡되어 나타나게 되는데, 이러한

결과를 방지하기 위하여 코퍼스 크기의 정규화(normalization)가 필요하다. 어휘다발의 정규화를 위한 '컷오프(cut-off)' 숫자를 확정해서 제시한 문헌은 존재하지 않지만, 선행연구에 의하면 코퍼스의 유형과 목적에 따라 구어 코퍼스의 경우 1백만 단어당 2-10회, 문어 코퍼스의 경우 1백만 단어당 20회[27][29][32][33], 25회[30][31], 또는 40회[25]로 다양하다.

선행 연구의 연구 목적과 연구 방법을 고려한 후, 본 연구에서는 1백만 단어당 20회 이상 비율로 출현하면서 최소 5개의 텍스트(논문)에 출현하는 어휘다발 선정의 기준으로 정하였다. 이때 본 연구에 사용된 두 개의 코퍼스 NSE와 BNC-BaAc는 그 크기가 다르기 때문에 정규화를 위해 NSE에서는 1백만 단어당 20회, 그리고 BNC-BaAc에서는 1백만 단어당 15회의 비율로 출현하는 어휘다발을 선정하였다. 이와 같은 어휘다발 선정에 관한 기준을 정한 후 어휘다발의 자동추출을 위해서는 AntConc 3.5.9[34]를 사용하였다. 자동으로 추출된 어휘다발은 발생 가능한 데이터 왜곡을 최소화하기 위해 몇 가지 조정과 세분화 처리 과정을 거쳤다. NSE와 BNC-BaAc는 각각 원자력과공학학이라는 특정 학문분야와 학술문으로 만들어진 코퍼스이므로 이 코퍼스에 불가피하게 포함된 해당 고유명사가 포함된 어휘다발을 제거함과 동시에 한 어휘다발의 뒷부분이 이어지는 어휘다발의 앞부분과 겹치는 경우를 합쳐서 빈도수의 왜곡을 최소화하였다. [표 2]는 이러한 정제과정을 거쳐 최종 선정된 NSE와 BNC-BaAc 어휘다발의 정보이다.

표 2. 어휘다발 타입 토큰 개수 (정제 후)

	NSE	BNC-BaAc
어휘다발 타입 개수	261	123
어휘다발 토큰 개수	10,218	3,533
타입/토큰 비율	2.53	3.48

3. 기능적 어휘다발 분류 기준

기능적 어휘다발은 크게 지시적(referential), 저자태도(stance), 그리고 담화조직(discourse organizing)과 같은 3개의 범주로 구분된다[30]. 이 세 개의 기능적

어휘다발은 그 하위 부류로 보다 구체적인 기능적 어휘다발로 부류로 구분된다. 지시적 어휘다발에는 시간표시 (time reference: *at the beginning of, at the end of, at the same time*), 장소표시 (place reference: *in the present work, on the inner surface, in the center of*), 기술 어휘다발 (descriptive: *the change in the, the location of the, the temperature of the*) 및 수량화 어휘다발 (quantification: *an increase in the, the magnitude of the, the size of the*)이 포함된다. 담화조직 어휘다발에는 비교/대비 (comparison/contrast: *a comparison of the, in comparison with, on the other hand*), 추론 (inferential: *are assumed to be, as a result of, on the basis of*), 주제초점 (focus: *it is necessary to, it is noted that, this means that the*) 및 구조화 (framing: *as a function of, in terms of the, with respect to the*) 어휘다발이 있다. 저자태도 어휘다발에는 비인칭형 인식과 가능성 (epistemic-impersonal, probable/possible: *can be applied to, is assumed that the, it is possible to*)의 기능을 하는 어휘다발과 기타 저자태도 기능의 어휘다발 (*be noted that the, considered in this study, was found to be*)이 포함된다. 마지막으로 위에서 설명된 어휘다발 기능의 범주에 포함되지 않는 기능을 하는 어휘다발이나 원자력과 핵공학 분야에 사용되는 어휘다발은 학문특징 범주로 분류하였다 (*of the fuel rod, pressurized water reactor power, the break flow rate*). 추출된 어휘다발은 어휘다발이 사용된 문맥을 파악하면서 그 기능을 분류하였다. 분류 과정에서 하나의 어휘다발이 두 가지의 기능으로 사용될 수도 있는데, 이러한 경우는 영어교육을 전공으로 하는 동료 연구자의 자문과 문맥상 보다 주도적으로 사용된 기능을 중심으로 분류하였다.

IV. 결과

1. 타입과 토큰의 분포

[그림 1]과 [그림 2]는 각각 NSE와 BNC-BaAc에서

사용된 전체 어휘다발에 대한 토큰(즉, 개별 어휘다발)과 타입(즉, 어휘다발 유형)의 분포와 빈도수를 보여준다.

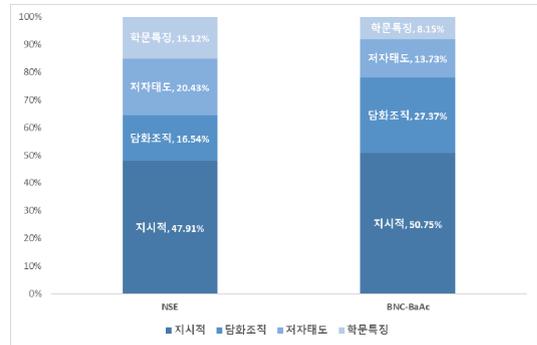


그림 1. NSE와 BNC-BaAc 기능 어휘다발 분포(토큰)

먼저 두 코퍼스의 토큰 분포에서 가장 눈에 띄는 점 중 하나는 NSE와 BNC-BaAc 공히 지시적 어휘다발이 가장 큰 분포로 각각 47.91%와 50.75%로 사용되었다. 그 다음으로는 NSE에서는 저자태도(20.43%)가, 그리고 BNC-BaAc에서는 담화조직(27.37%)이 큰 분포로 사용되었다. NSE와 BNC-BaAc에 사용된 어휘다발 타입을 살펴보면 여전히 지시적 어휘다발이 가장 큰 분포로 사용되었지만, BNC-BaAc에서 담화조직의 분포가 27.37%로 약간 늘어난 반면, NSE에서는 저자태도가 15.71%로 조금 줄어들었다.

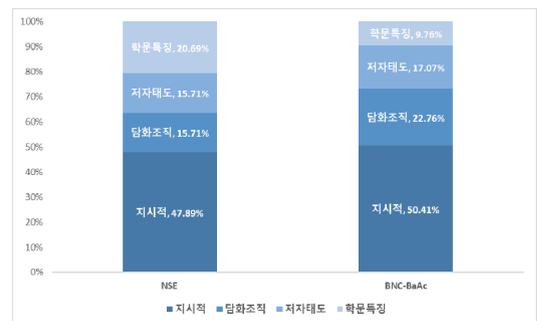


그림 2. NSE와 BNC-BaAc 기능 어휘다발 분포(타입)

2. 기능적 어휘다발 사용의 다양성 비교

NSE와 BNC-BaAc에 사용된 어휘다발 사용의 다양성 측정하기 위해 각 코퍼스에 따라 타입/토큰 비율을 계산하였다. [표 3]에 의하면, 코퍼스 전체에서 NSE는

2.55, 그리고 BNC-BaAc 4.16의 타입/토큰 비율을 나타내는데, 이는 BNC-BaAc에 조금 더 다양한 어휘다발이 사용되었다는 것을 의미한다. 각각의 기능적 어휘다발의 영역을 살펴볼 때 NSE에서 지시적 어휘다발(2.55)과 저자태도(1.96)와 관련된 어휘다발 사용이 BNC-BaAc에 비해 상대적으로 획일화된 유형으로 사용되었다는 점이다.

표 3. NSE와 BNC-BaAc 기능적 어휘다발의 타입/토큰 비율

	NSE	BNC-BaAc
지시적	2.55	3.46
담화조직	2.43	2.90
저자태도	1.96	4.33
학문특징	3.50	3.82
코퍼스 전체	2.55	4.17

3. 기능적 어휘다발의 부류와 유형 비교

[표 4]는 각 범주의 어휘다발 분포의 비율을 보여준다.

표 4. NSE와 BNC-BaAc의 어휘다발 부류 분포(토큰-타입)

부류	유형	NSE	BNC-BaAc	어휘다발의 예
지시적	시간표시	2.32% (237)	4.00% (398)	<i>at the same time</i>
		2.30% (6)	10.83% (13)	
	장소표시	13.42% (1,371)	0.00% (0)	<i>in the present work</i>
		13.41% (35)	0.00% (0)	
	기술	11.33% (1,158)	6.86% (233)	<i>the location of the</i>
		12.64% (33)	9.17% (11)	
	수량화	8.26% (844)	7.10% (241)	<i>the size of the</i>
		9.95% (26)	6.67% (8)	
	구조화	12.58% (1,285)	27.44%(921)	<i>in terms of the</i>
		9.58% (25)	25.00% (30)	
담화조직	비교/대조	4.28% (437)	3.77% (128)	<i>on the other hand</i>
		3.07% (8)	1.67% (2)	
	추론	3.05% (312)	5.80% (197)	<i>as a result of</i>
		2.68% (7)	4.17% (5)	
	주제초점	9.21% (941)	15.14% (514)	<i>it is needed that</i>
		9.96% (26)	16.67% (20)	
저자태도	인식적 기능	8.44% (862)	9.81% (333)	<i>can be applied to</i>
		9.20% (24)	10.83% (13)	
	기타	12.39% (1,226)	4.48% (152)	<i>was found to be</i>
		6.51% (17)	5.83% (7)	
학문특징		15.12% (1,545)	8.48% (288)	<i>the break flow rate</i>
		20.69% (54)	9.17% (11)	

계	100% (10,218)	100% (3,396)	
	100% (261)	100% (120)	

주) 각 코퍼스 열의 셀에서 상단 값은 토큰, 하단 값은 타입의 정보임. 괄호 안의 숫자는 빈도수임. 어휘다발의 예는 NSE에서 추출함.

NSE와 BNC-BaAc에 나타난 지시적 어휘다발은 비슷한 분포 비율을 보여주지만[그림 2], 어휘다발에 포함되어있는 세부적인 기능 분류에 있어서는 두 코퍼스에 서로 다른 양상을 보여준다. 예를 들어, NSE에는 장소 표시 어휘다발은 35개의 타입(13.41%)이 나타났지만, BNC-BaAc에는 그러한 어휘다발 타입은 나타나지 않았다. 이와는 대조적으로 BNC-BaAc에서는 시간표시(10.83%)와 구조화(25.00%) 어휘다발 타입의 비율은 각각 NSE의 시간표시(2.30%)와 구조화(9.58%) 어휘다발 타입의 비율보다 높았다. 또한 담화조직 어휘다발의 범주에도 눈에 띄는 대조가 포착되었다. 비교/대조, 추론, 주제초점의 세 가지 패턴 중 BNC-BaAc의 주제초점(16.67%) 비율이 NSE(9.96%)보다 높다. NSE와 BNC-BaAc에서 저자태도 어휘다발 사용의 차이는 미미하지만 NSE에서 분야별 어휘다발의 비율은 20.69%로 BNC-BaAc보다 높았다. 이러한 기능적 어휘다발의 분포는 NSE가 원자력과공학학 분야의 학술논문으로 구성되었기 때문에 나타난다고 예상할 수 있고, 결과적으로 이 분야에서 사용되는 어휘다발의 기능이 일반적인 학문 분야에 사용되는 어휘다발과는 구별되는 특징으로 나타난다는 점을 보여준다.

4. 기능적 어휘다발의 분포 차이의 기여 요인

[표 5]는 NSE와 BNC-BaAc에 사용된 기능적 어휘다발 토큰의 분포를 비교한 결과이다.

표 5. 기능적 어휘다발 분포에 대한 카이제곱 분할표와 표준화 잔차(토큰)

		지시적	담화조직	저자태도	학문특징
$\chi^2 = 11.345, df = 3, p = 0.000, \text{크레머 계수 } V = 0.145,$					
NSE	관찰빈도	4,895	1,690	2,088	1,545
	기대빈도	4,986.35	1,964.68	1,918.34	1,366.62
	표준화잔차	-1.29	-5.82	3.87	4.83
BNC-BaAc	관찰빈도	1,793	921	485	288

	기대빈도	1,701.65	664.32	654.66	466.38
	표준화잔차	2.21	9.96	-6.63	-8.26

주) 효과크기: 0.15(small effect size); $p < 0.01$.

[표 5]에 의하면 NSE와 BNC-BaAc에 나타난 기능적 어휘다발 타입의 분포를 카이-제곱 검정으로 분석한 결과 두 코퍼스에 사용된 기능적 어휘다발의 사용 분포가 0.000 유의수준에서 유의미한 차이로 다르다는 결과를 얻었다($\chi^2 = 7.815$, $df = 3$, $p = 0.000$, 크레머계수 $V = 0.145$). 이때 어휘다발의 네 가지 기능 중 어떤 기능의 어휘다발의 사용이 이 유의미한 차이에 대한 기여 요인(contributor)으로 작용했는지 확인해 볼 필요가 있다. 이를 위해 표준화 잔차(standardized residual)를 사용했다. 표준화 잔차는 관찰 빈도와 기대 빈도를 비교하는 비율로 정해지는데, 이는 카이제곱 검정을 통하여 얻어지는 추정 회귀식(표준편차)으로부터 관찰 빈도가 얼마나 떨어져 있는지를 나타낸다. 일반적으로 표준화 잔차의 절대값이 1.96보다 크다면 이상 점으로 간주하는데, 이러한 차이를 나타내는 범주가 카이검증을 통한 통계적 차이에 기여 요인으로 작용한다고 해석한다. 즉, 표준화 잔차 값이 1.96보다 크다는 것은 셀의 관찰 빈도가 기대 빈도 보다 크게 다르다는 것을 의미한다. 표준화 잔차값의 앞에 붙는 음수(-)의 부호 관찰 빈도가 예상 빈도보다 높거나 낮을 경우를 나타낸다[30].

NSE와 BNC-BaAc의 기능적 어휘다발 토큰 사용에 대한 유의미한 차이는 만드는 기여 요인은 NSE에 사용된 저자태도와 학문적 특징(즉, 원자력과학공학) 어휘다발 토큰이 통계적 예상 빈도보다 높게 사용되었다는 것이다(표준화잔차 3.82). 원자력과학공학의 학술 논문으로 이루어진 NSE에서는 누구든지 예상할 수 있듯이 관련 학문의 전문용어 및 개념이 포함된 어휘다발이 일반 학술논문으로 만들어진 BNC-BaAc에 비해 많이 사용될 수 있을 것이다. 또 저자태도 어휘다발 토큰은 *is assumed that the, is assumed to be, is considered to be, is defined as the, is well known that*와 같이 저자의 인식론적 관점에 대한 어휘다발과 *can be applied to, can be classified into, can be considered as, can be expressed as, can be found in, can be obtained by*와 같이 저자

가 인식하는 불확실성/확실성과 관련한 어휘다발이 다수 사용되었다.

이와는 별도로 NSE에는 담화조직 어휘다발이 통계적 예상 빈도보다 낮게 사용되었다(표준화잔차-5.82). 담화조직 어휘다발 중 추론 어휘다발인 *as a result of, as a result the, on the basis of*나 비교/대조 어휘다발인 *on the other hand (the), than that of the, the comparison of the, the difference between the*와 같은 어휘다발이 자주 사용되었다. 이와 더불어 *for the analysis of, for the calculation of, for the case of, for the design of*와 같은 주제초점 어휘다발이 담화조직 어휘다발에 포함되어있는데, 이러한 담화조직 이공계 학술논문, 특히 원자력과학공학 학술논문의 논문 구성과 관련한 특성을 반영하기 위한 어휘다발로 여겨진다.

이때 어휘다발의 사용 개수 비교와는 다르게 NSE와 BNC-BaAc에 얼마나 '다양한' 어휘다발이 각각의 코퍼스에 서로 사용되었는지 비교해 볼 필요가 있다. 표5는 NSE와 BNC-BaAc에 사용된 기능적 어휘다발 타입의 분포 비교 결과이다. 그 결과에 의하면 NSE와 BNC-BaAc에 나타난 기능적 어휘다발 타입의 분포를 카이-제곱 검정으로 분석한 결과 두 코퍼스에 사용된 기능적 어휘다발의 사용 분포가 0.05 유의수준에서 유의미한 차이로 다르다는 결과를 얻었다($\chi^2 = 7.815$, $df = 3$, $p = 0.026$, 크레머계수 $V = 0.145$). 이러한 통계적 유의미성에 대한 기여 요인은 BNC-BaAc에 학문적 특징을 나타내는 기능적 어휘다발이 통계적 기대치보다 적게 나타난 것이다(표준화잔차 -2.11).

표 6. 기능적 어휘다발 분포에 대한 카이제곱 분할표와 표준화 잔차(타입)

		$\chi^2 = 7.815$, $df = 3$, $p = 0.026$, 크레머계수 $V = 0.156$,			
		지시적	담화조직	저자태도	학문특징
NSE	관찰빈도	125	41	41	54
	기대빈도	127.77	47.14	41.69	44.41
	표준화잔차	-0.24	-0.89	-0.11	1.44
BNC-BaAc	관찰빈도	62	28	20	11
	기대빈도	59.23	21.86	19.32	20.59
	표준화잔차	0.36	1.31	0.15	-2.11

주) 효과크기: 0.16 (small effect size); $p < 0.05$.

어휘다발 토큰 분포 비교 결과인 [표 5]와 어휘다발 타입 분포 비교 결과인 [표 6]의 결과를 교차 검토해 볼 때 기능적 어휘다발 타입과 토큰의 사용은 NSE와 BNC-BaAc에 각각 유의미하게 다른 사용 분포를 보여 주는데, 이러한 결과가 시사하는 점은 다음과 같다. 첫째, NSE와 BNC-BaAc에 있어서 기능적 어휘다발은 서로 통계적으로 유의미한 분포의 차이로 사용된다. 둘째, NSE와 BNC-BaAc의 차이의 기여 요인으로는 학문의 특징을 나타내는 어휘다발 사용의 차이로, 즉 원자력과공학학 학술논문의 특징적인 어휘다발과 일반 학술문의 어휘다발의 사용의 차이로 볼 수 있다. 이와 더불어 기능적 어휘다발의 토큰과 타입 사용분포의 비교 결과는 그 기여 요인에 있어서 원자력과공학학 학술문에 나타나는 어휘다발 사용의 특징을 보여준다. 즉, 어휘다발 타입의 사용에서는 원자력과공학학의 특징은 학문적 특징을 드러내는 어휘다발 사용이 중요한 기여 요인이었다. 하지만, 일반 학술문과의 어휘다발 토큰의 비교에서는 원자력과공학학 학술문은 저자태도 어휘다발이 통계적인 기대치보다 유의미한 크기로 더 많이 사용된 반면, 담화조직 어휘다발은 통계적 기대치보다 적게 사용된 것으로 나타났다. 우연이지만 실제로, NSE에서 담화조직과 저자태도는 어휘다발은 동일하게 41 타입이 추출되었다. 이 어휘다발을 토큰의 수라는 관점에서 비교해 볼 때 원자력과공학학 논문에서는 41개의 저자태도 어휘다발을 매우 활발하고 반복적으로 사용된 반면(2,088개의 토큰), 담화조직 어휘다발 41개는 통계적 기대치보다는 덜 사용되었다(1,690개의 토큰). 어휘다발 타입의 실제 사용 용례인 토큰의 개수로 볼 때 원자력과공학학 학술문에 저자태도를 표현함에 있어서 한정된 숫자의 어휘다발을 반복적으로 그리고 집중적으로 사용한다는 사실을 알 수 있다. 반면 담화조직 어휘다발은 일반 학술문에서 사용된 것만큼 사용되지 않는 것으로 나타났다.

V. 결론

본 연구는 원자력과공학학 영어 학술 논문과 일반 학술 논문에 나타나는 기능적 어휘다발의 분포의 차이가 있는지, 있다면 어떤 기능적 어휘다발 부류와 유형이

그 차이에 대한 기여 요인으로 작용했는지 분석하였다. 이를 통해 얻어진 결론은 다음과 같다.

첫째, 원자력과공학학과 일반 학술 논문 모두 지시적 기능을 담당하는 어휘다발이 가장 많이 사용되었다. 지시적 어휘다발이 다수 출현하는 것은 저자태도나 담화조직이 상대적으로 많이 사용되는 대화문과 비교하였을 때 학술문에 나타나는 두드러진 특징으로 볼 수 있다[26]. 그러나 원자력과공학학 논문에서는 저자태도와 담화조직 어휘다발의 순으로, 일반 논문에서는 담화조직과 저자태도의 순으로 어휘다발의 기능이 분포되었는데, 이는 일반 학술논문과 비교될 수 있는 원자력과공학학 논문에 사용되는 기능적 어휘다발의 특징으로 볼 수 있다.

둘째, 두 학문 분야에 사용된 기능적 어휘다발의 다양성을 타입/토큰 비율을 통해 알아본 결과 원자력과공학학 학술 논문에서는 일반 학술 논문과 비교할 때 상대적으로 다양하지 않은 어휘다발이 사용된다는 사실이 밝혀졌다. 특히, 원자력과공학학 학술 논문에서는 저자태도 어휘다발이 상대적으로 해당 학술문 내의 지시적, 담화조직 어휘다발과 비교했을 때, 그리고 일반 학술 논문의 여타 기능적 어휘다발과 비교했을 때 다채롭지 못한 사용 분포를 나타냈다.

셋째, 원자력과공학학 학술 논문에 사용된 기능적 어휘다발의 세 가지 부류 있어서, 일반 학술 논문에 비해 지시적 어휘다발의 장소표시 어휘다발이 많이 사용된 반면, 담화조직 어휘다발에는 주체초점 어휘다발이 적게 사용되었고, 원자력과공학학 학술 논문에는 저자태도 어휘다발은 그 차이가 미미하였다. 이와 더불어 원자력과공학학의 학문적 특징을 나타내는 어휘다발은 상당한 대조를 보이며 다수 사용되었는데, 실제 사용된 어휘다발의 토큰 빈도를 분석한 결과 이러한 차이의 기여 요인은 원자력과공학학에서는 담화조직 어휘다발 사용이 적었다. 이는 응용언어학이나 경제학과 같은 인문사회계열의 학문에 비해 전기공학이나 생물학과 같은 이공계 학문의 학술 논문에서 상대적으로 담화조직 어휘다발이 적게 사용된다는 연구 결과와도 부합한다[27]. 또한 저자태도와 학문적 특징을 나타내는 어휘다발이 통계적으로 유의미하게 많이 사용된다는 사실이 밝혀졌는데, 이는 저자태도 어휘다발 사용은 역사 분야

와 비교했을 때 생물학 분야에서 저자태도, 특히 인식론과 관련한 어휘다발 사용이 상대적으로 높다는 기존의 연구와도 부합하는 결과이다[28]. 그럼에도 불구하고 사용된 어휘다발의 타입을 비교한 결과 유의미한 차이에 대한 기여요인이 나타나지 않아 원자력과화학공학 학술 논문에서는 한정된 수의 어휘다발이 반복되어 사용된다는 사실도 밝혀졌다.

최근 인문사회 과학기술분야를 막론하고 학문의 발전과 교류가 이전보다 더 빠르게 진행되고 있다. 이렇게 급속한 진행에 국제 학계에서 영어의 역할은 이전보다 더 강조되고 있고, 학술목적영어에 관한 연구와 교육은 이전보다 더 전문적이고 세분화 되어야 한다. 본 연구에서 밝혀진 원자력과화학공학 학술 논문 영어에 나타난 기능적 어휘다발의 특징은 해당 분야에서 사용되는 다른 분야의 영어 사용과 구별되는 특징의 단면을 보여주었다. 이러한 측면에서 볼 때 본 연구의 결과로 얻어진 기능적 어휘다발의 특징은 추후 원자력과화학공학 학술목적영어 교육의 기초자료로 사용될 것으로 기대한다. 또한 기존 학술목적영어 교육의 보다 진정성과 실재성을 담보하여 교육의 효율성을 높이기 위해 코퍼스를 활용한 관련 분야에 대한 영어 분석이 보다 다양한 방향으로 연구될 필요가 있다.

참 고 문 헌

- [1] M. Charles, "English for academic purposes," In B. Paltridge and S. Starfield (Eds.), *The Handbook of English for Specific Purposes*, pp.137-153, Wiley-Blackwell, 2013.
- [2] A. Mauranen, "English as the lingua franca of the academic world," In D. Belcher, A. Johns and B. Paltridge (Eds.), *New Directions in English for Specific Purposes Research*, pp.94-117, The University of Michigan Press, 2011.
- [3] K. Conklin and N. Schmitt, "Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers?," *Applied Linguistics*, Vol.29, No.1, pp.72-89, 2008.
- [4] J. Flowerdew and M. Peacock, *Research perspectives on English for academic purposes*, Cambridge University Press, 2001.
- [5] K. Hyland, *Second Language Writing*, Cambridge University Press, 2009.
- [6] K. Hyland, "Bundles in academic discourse," *Annual Review of Applied Linguistics*, Vol.32, pp.150-169, 2012.
- [7] A. Johns, "Tertiary undergraduate EAP: Problems and possibilities," In. D. Belcher (Ed.). *English for specific purposes in theory and practice*, pp.41-59, The University of Michigan Press, 2009.
- [8] R. R. Jordan, *English for Academic Purposes*, Cambridge, Cambridge University Press, 1997.
- [9] H. Nesi and S. Gardner, *Genres across the Disciplines: Student Writing in Higher Education*, Cambridge University Press, 2012.
- [10] J. Swales, *Genre Analysis: English in Academic and Research Settings*, Cambridge University Press, 1990.
- [11] J. Swales, *Research genres: Exploration and applications*, Cambridge University Press, 2004.
- [12] D. Belcher, *English for Specific Purposes in Theory and Practice*, University of Michigan Press, 2009.
- [13] D. Belcher, A. M. Johns, and B. Paltridge (Eds.), *New Directions in English for Specific Purposes Research*, University of Michigan Press, 2011.
- [14] T. Dudley-Evans and M. J. St. John, *Developments in ESP: A Multi-Disciplinary Approach*, Cambridge University Press, 1998.
- [15] L. Grabowski, "Keywords and lexical bundles within pharmaceutical discourse," *English for Specific Purposes*, Vol.38, pp.23-33, 2015.
- [16] H. Nesi and H. Basturkmen, "Lexical bundles and discourse signalling in academic lectures," *International Journal of Corpus Linguistics*, Vol.11, No.3, pp.283-304, 2006.
- [17] B. Paltridge and S. Starfield (Eds.), *The Handbook of English for Specific Purposes*,

- Wiley-Blackwell, 2013.
- [18] D. J. Salazar, *Biomedical English: A Corpus-Based Approach*, John Benjamins, 2013.
- [19] 권미분, “관광학술지 영문초록 코퍼스의 어휘다발에 관한 연구,” 언어과학, 제26권, 제3호, pp.1-19, 2019.
- [20] 이성민, 하명호, “화학공학영어 코퍼스기반 키워드 및 어휘다발 분석: 교재와 논문 코퍼스를 중심으로,” 영어영문학연구, 제59권, 제1호, pp.271-290, 2017.
- [21] 최인영, “코퍼스(Corpus) 기반 항공안전영어 어휘 리스트 구축: IATA의 Safety Report를 중심으로,” 관광레저연구, 제32권, 제12호, pp.299-316, 2020.
- [22] J. Y. Lee and H. J. Lee, “Lexical bundles in computer science research articles: A corpus-based study,” *International Journal of Contents*, Vol.14, No.4, pp.70-75, 2018.
- [23] D. Biber, S. Conrad, and V. Cortes, “If you look at...: Lexical bundles in university teaching and textbooks,” *Applied Linguistics*, Vol.25, pp.371-405, 2004.
- [24] S. Conrad and D. Biber, “The frequency and use of lexical bundles in conversation and academic prose,” *Lexicographica*, Vol.20, pp.56-71, 2004.
- [25] D. Biber and F. Barbieri, “Lexical bundles in university spoken and written registers,” *English for Specific Purposes*, Vol.26, No.3, pp.263-286, 2007.
- [26] D. Biber, S. Johansson, G. Leech, S. Conrad, and E. Finegan, *Longman Grammar of Spoken and Written English*, Longman, 1999.
- [27] K. Hyland, “As can be seen: Lexical bundles and disciplinary variation,” *English for Specific Purposes*, Vol.27, No.1, pp.4-21, 2008.
- [28] V. Cortes, “Lexical bundles in published and student disciplinary writing: Examples from history and biology,” *English for Specific Purposes*, Vol.23, No.4, pp.397-423, 2004.
- [29] H. Kučera and W. N. Francis, *Computational Analysis of Present-Day American English*, Brown University, 1967.
- [30] Y. H. Chen and P. Baker, “Lexical bundles in L1 and L2 academic writing,” *Language Learning and Technology*, Vol.14, No.2, pp.30-49, 2010.
- [31] D. Nam, “Functional distribution of lexical bundle in a native and nonnative students’ argumentative writing,” *Journal of Asia TEFL*, Vol.14, No.4, pp.703-716, 2017.
- [32] R. Reppen, “Exploring L1 and L2 writing development through collocations: A corpus based look,” In A Barfield & H. Gyllstad (Eds.), *Researching Collocations in Another Language*, Palgrave Macmillan, pp.49-59, 2009.
- [33] Y. Wei and L. Lei, “Lexical bundles in the academic writing of advanced Chinese EFL learners,” *RELC Journal*, Vol.42, pp.155-165, 2011.
- [34] L. Anthony, AntConc (Version 3.5.9) [Computer Software]. Waseda University. Available from <http://www.antlab.sci.waseda.ac.jp>, 2020.

저 자 소 개

남 대 현(Daeheon Nam)

정희원



- 1996년 2월 : 한국외국어대학교 노어과(문학사)
- 1999년 2월 : 한국외국어대학교 노어노문학과(문학석사, 어학 전공)
- 2004년 5월 : Indiana University, Bloomington (문학석사, 전산언어학 전공)

- 2010년 6월 : Indiana University, Bloomington(언어교육학 박사, 영어교육 전공)
- 2011년 3월 ~ 2021년 2월 : 울산과학기술원 조교수
- 2021년 3월 ~ 현재 : 사이버한국외국어대학교 부교수 <관심분야> : 코퍼스기반 영어교육, 코퍼스 언어학, 네트워크 분석