

12조 최종보고서

문/이과 및 계열 구분에 따른 논문 작성 방식 비교

박찬희, 이다은, 장윤나, 전순표

I. 서론

이 연구는 약 500만 어절 규모의 분야별 한국어 논문 자료를 수집, 전산화하여 분석하는 것을 목표로 한다. 분석도구로는 'AntConc'의 concordance, N-grams, word list 등이 사용되며 시각화를 위해 'Pajek'을 사용하였다. 연구의 목적은 격식적, 학문적 글쓰기의 정점에 있는 논문의 어휘, 문장구조를 분석하고 각 분야별 논문의 글쓰기 방식의 차이를 확인하는 것이다.

논문의 언어 사용 양상 분석을 통해 얻을 수 있는 효과는 크게 외국어 화자에 대한 것과 모국어 화자에 대한 것으로 나눌 수 있다. 먼저 외국어로서 한국어를 배우는 학습자에게 있어서 학문적 텍스트를 읽고 쓰는 것은 매우 어려운 일이다. 이러한 자료들이 어렵게 느껴지는 이유는 사용되는 어휘와 표현이 일상적인 것과 크게 다르기 때문일 것이다. 본 연구를 통해 전산화된 자료는 학문적 텍스트에 대한 실제적 교육 자료가 될 수 있을 것이다. 모국어 화자에게 있어도 논문은 어려운 텍스트에 속하는데, 이는 다른 담화 유형에 비해 우리가 논문을 비교적 늦게 접하기 때문일 것이다. 모국어 화자는 연구 결과를 토대로 다른 격식적인 텍스트나 학문적인 텍스트와 비교했을 때 논문이 가지는 특징이 무엇인지 세밀하게 분석해보고 특히 분야별 논문의 제목, 서론, 결론 등의 형식적 특징과 핵심 요소들을 파악하여 이를 논문의 생산과 수용에 활용할 수 있을 것이다.

본 연구에서는 구체적으로 분야별 논문의 특징적인 단어나 문장 구조를 추출하고 비교해 볼 것이다. 분야별 글쓰기 방식에 차이가 있다는(또는 그래야만 한다는) 주장은 오래전부터 논의 되었으며 특히 이공계열, 인문계열 간의 차이가 중요한 것으로 여겨지고 있다. 이에 따라 실제로 많은 글쓰기 수업에서 전공 분야에 맞는 글쓰기 수업을 진행하고 있기도 하다. 본 연구에서는 학문 분야별 특성 파악과 더불어, 구축한 코퍼스를 다시 문/이과의 계열별로 분류한 뒤 공통점과 차이점을 찾아내고자 한다. 이를 통해 분야별 특성에 맞는 글쓰기 수업 방식이 강화되거나 보완되는 효과를 거둘 수 있을 것이다. 이를 위해 계열별, 전공별로 전체 상위빈도 어휘, N-gram, 품사별 상위빈도 어휘와 논문 제목의 형태론적 특징을 추출해보고 이것이 각 분야의 특성을 어떻게 반영하고 있는지 연구해보고자 한다. 또한 양적 연구로는 놓치기 쉬운 논문의 구조, 제목의 통사적 특징을 알아보기 위해 각 전공별 10개 논문의 서론, 결론의 구조 분석과 제목의 통사적 특징에 대한 연구도 포함할 것이다. 연구를 통해 분야별 논문 특성의 일반화의 가능성을 찾아볼 수 있기를 기대한다.

II. 코퍼스 구축 과정

코퍼스 구축을 위해 먼저 kiss(한국학술정보)의 주제 분류를 참고하여 계열별로 논문 300개씩을 수집하였다. 6개의 계열 중 어문학, 인문과학, 사회과학은 문과로, 자연과학, 공학, 의학은 이과로 묶을 수 있다. 이 외에 kiss는 농학, 수해양, 예체능도 분류하였으나 이들은 문과나 이과로 특정하기 어려운 점을 감안하여 제외했다. 구체적인 전공명은 [그림 1]과 같다. 이때 어문학, 인문과학, 공학, 의학은 각자 말았으며, 사회과학과 자연과학의 경우 두 명이 나누어 논문을 모았다.

문과	인문과학	한국사, 동양사, 서양사, 인류학, 동양철학, 서양철학, 종교학, 민속학
	사회과학	법학, 정치/외교학, 행정학, 경제학, 경영학, 회계학, 지역개발, 관광학, 지역학, 사회학, 사회복지, 신문방송, 심리학, 문헌정보학, 교육학, 인문지리
	어문학	언어학, 국어국문학, 중어중문학, 일어일문학, 아랍어문학, 영어영문학, 불어불문학, 독어독문학, 서어서문학, 노어노문학
이과	자연과학	수학, 통계학, 물리학, 화학, 생물, 지질, 자연지리, 천문학, 대기화학
	공학	기계공학, 조선공학, 항공공학, 산업공학, 전기공학, 전자공학, 화학공학, 환경공학, 자원공학, 토목공학, 건축공학, 컴퓨터공학
	의약학	해부학, 병리학, 보건학, 내과학, 소아과학, 피부과학, 신경과학, 정신과학, 외과학, 산부인과학, 정형외과학, 신경외과학, 안과학, 마취과학, 방사선과학, 재활의학, 가정의학, 치의학, 간호학, 약화학, 한의학

그림 1 : 세부 전공 구분

계열별 총 논문의 개수는 300개로 하되, 계열 내에서 특정 전공에 편중되지 않도록 세부 전공에서 비슷한 비율로 논문을 수집하였다. 예를 들어, 공학 계열의 경우 건설·토목공학, 기계공학, 산업공학, 전기·전자공학, 조선·항공공학, 컴퓨터공학, 화학공학, 환경공학이라는 8개의 세부 전공에서 각각 35~40개의 논문을 코퍼스 구축에 활용하였다. 또한 시대에 따라 논문 작성 방식이 차이가 날 수 있기 때문에 모든 논문은 2000년대에 작성된 것으로 한정했다. 이는 연구목적이 최근의 문·이과 논문 작성 방식에 유의미한 차이가 있는지를 살펴보는 것임을 고려한 것이다. 이렇게 수집한 논문 pdf 파일을 하나의 파일로 병합한 후 txt 파일로 변환하였다.

논문 수집을 마친 뒤에는 전처리 작업에 들어갔다. 코퍼스 분석의 정확성을 높이기 위해 논문에 포함되어 있는 초록, 참고문헌, 서지 정보 등의 불필요한 정보들을 제거하고자 했다. 또한 텍스트 파일로 변환했을 때, 텍스트의 중간에 그림 및 표, 저자 및 발행기관, 각주 등이 삽입된 경우가 많았다. 최대한 이를 제거하여 코퍼스 내 단어의 형태가 온전히 유지되도록 했다. 전처리는 각자 본인이 수집한 논문들을 맡아서 진행했으며, R이나 파이썬을 활용한 후 처리되지 않은 부분은 메모장 등에서 직접 수작업하여 텍스트를 정제했다. 그런데 이 작업 후에 각자 전처리 방식에 차이가 있어 계열별 분석 결과에 통일성이 다소 떨어지는 측면이 있음을 확인하였다. 그래서 괄호 내 숫자 삭제, 영문은 전체 삭제, 특수문자 삭제, 3개 이상의 숫자 삭제 등의 공통적인 기준을 정하고 다시 처리하였다. 이때 영문과 숫자를 일괄적으로 삭제한 이유는 본 연구가 '한국어'로 된 논문에서 문·이과, 계열별 차이가 어떻게 나타나는지 밝히는 데에 초점이 맞춰져 있기 때문이다. 이과 논문에서 수식이나 학술 용어 때문에 영문과 숫자가 훨씬 많이 쓰이는 것은 사실이다. 하지만 영문과 숫자를 삭제하지 않을 경우, 이들이 고빈도로 집계되어 오히려 분석 과정에서 혼란을 야기하고 정확성을 떨어뜨릴 수 있다. 결국 '한국어'로 이들을 어떻게 표현하는가에 관련한 연구이므로 영문과 숫자는 고려하지 않았음을 밝히는 바이다. 이렇게 수집된 분야별 및 문이과별 코퍼스의 총 어휘 수와 어절 수는 [표 1]와 같다. 한편 어휘 수의 경우, 계열별로 중복이 발생하기 때문에 합계와 부분의 합이 일치하지 않아 제거하였다.

분야	어휘 수(type)	어절 수(token)
문과	484,575	3,390,023
이과	193,732	1,662,741
자연과학	75,520	551,134
공학	73,008	378,852
의약학	99,346	732,755
인문과학	221,060	1,133,023
사회과학	189,026	1,280,955
어문학	189,164	1,089,147
합계	x	5,165,866

표 1 문/이과 및 계열별 어휘 및 어절 수

III. 코퍼스 분석

1. 초기 분석 방식

(1) 상위 200개 고빈도 단어 목록 추출 – AntConc

각 분야별 코퍼스에서 자주 쓰이는 표현들을 분석하기 위하여 AntConc의 wordlist 추출 기능을 사용하였다.

인문과학	사회과학	어문학	자연과학	공학	의약학
1 10453 0.923% 수	13402 1.046% 수	9818 0.901% 수	6037 1.095% 수	4207 1.110% 있다	5840 0.797% 및
2 8762 0.773% 있다	12638 0.987% 있다	9677 0.888% 있다	4613 0.837% 있다	3896 1.028% 수	5604 0.765% 수
3 6978 0.616% 대한	10139 0.792% 대한	9361 0.859% 의	4024 0.730% 대한	2888 0.762% 및	4896 0.668% 대한
4 6193 0.547% 있는	6776 0.529% 있는	6985 0.641% 이	3685 0.669% 의	2522 0.666% 의	4752 0.649% 있다
5 5508 0.486% 이	5963 0.466% 및	5040 0.463% 그	2446 0.444% 는	1800 0.475% 대한	4425 0.604% 본
6 5484 0.484% 그	4700 0.367% 년	4767 0.438% 있는	2434 0.442% 및	1735 0.458% 있는	3527 0.481% 있는
7 4044 0.357% 것이다	4682 0.366% 것이다	4725 0.434% 대한	2386 0.433% 있는	1583 0.418% 는	3431 0.468% 의
8 3109 0.274% 이러한	4567 0.357% 것으로	4712 0.433% 는	2299 0.417% 이	1444 0.381% 를	3292 0.449% 후
9 3021 0.267% 한다	4270 0.333% 이러한	4293 0.394% 것이다	1999 0.363% 수학	1374 0.363% 본	3259 0.445% 것으로
10 2955 0.261% 것은	3910 0.305% 이	4011 0.368% 가	1926 0.349% 를	1273 0.336% 예	2136 0.292% 로
11 2837 0.250% 것이	3762 0.294% 그	3483 0.320% 와	1841 0.334% 예	1268 0.335% 그림	1970 0.269% 를
12 2798 0.247% 하는	3649 0.285% 의	3324 0.305% 예	1775 0.322% 와	1245 0.329% 것으로	1931 0.264% 경우
13 2765 0.244% 년	3639 0.284% 영향을	3134 0.288% 를	1759 0.319% 본	1213 0.320% 경우	1925 0.263% 이
14 2594 0.229% 할	3481 0.272% 본	2930 0.269% 은	1701 0.309% 것으로	1178 0.311% 이	1877 0.256% 또한
15 2563 0.226% 통해	3295 0.257% 통해	2861 0.263% 을	1611 0.292% 은	1153 0.304% 통해	1821 0.249% 연구
16 2381 0.210% 그리고	3031 0.237% 또한	2567 0.236% 한다	1581 0.287% 과	1136 0.300% 와	1803 0.246% 명
17 2362 0.208% 아니라	2960 0.231% 위한	2538 0.233% 이러한	1506 0.273% 에서	1012 0.267% 또한	1795 0.245% 영향을
18 2271 0.200% 다른	2929 0.229% 한다	2442 0.224% 한	1441 0.261% 을	994 0.262% 은	1791 0.244% 중
19 2250 0.199% 의	2773 0.216% 알	2431 0.223% 에서	1421 0.258% 수학적	979 0.258% 위한	1671 0.228% 위한
20 2245 0.198% 것으로	2754 0.215% 하는	2404 0.221% 과	1379 0.250% 통해	978 0.258% 위해	1659 0.226% 년

그림 2 : wordlist기능으로 추출한 상위 20개 단어

[그림 2]는 빈도수 순위에 따른 단어의 출현 빈도, 전체 코퍼스 대비 비율, 그리고 해당 단어 순으로 상위 20개 단어들을 나열하고 있다. 분야마다 300개의 논문으로 구성된 코퍼스 6개 전체를 일대일로 대조하는 데 현실적으로 어려움이 따를 것이라고 판단하여 이러한 방식을 채택하였다. 논문별 길이가 일정하지 않다는 점은 단어의 출현 빈도에도 고스란히 적용되어 드러났다. 즉, 사회과학과 인문과학 분야의 논문 길이가 가장 길었으며 상대적으로 자연과학, 의약학 분야는 짧은 편이었다. 이러한 한계를 고려하여 절대적 빈도수와 함께 상대적 빈도수를 감안하였다.

흥미롭게도 계열별로 공통적으로 드러나는 표현이 상위 단어들의 상당수를 차지하고 있었다. 위 그림에 제시된 상위 20개의 단어는 물론, 상위 200개 고빈도 단어 목록에서 뚜렷한 차이를 찾기는 어려웠다. 이후 Pajek, N그램 추출과 같은 다양한 접근 방법을 통해 이러한 문제점을 해결하고자 하였다.

그림 4 : 상위 50개 분석도

마찬가지로 상위 50개 공기어들을 시각화한 [그림 4]에서 6개 분야 모두의 공기어로 채택된 키워드는 다음과 같다.

‘있다’, ‘수’, ‘의’, ‘것으로’, ‘있는’, ‘이’, ‘대한’.
‘통해’, ‘는’, ‘이러한’, ‘경우’, ‘또한’, ‘그’, ‘본’, ‘및’

기존 7개 공기어에서 8개가 추가되었다. 상위 20개에서 연결 정도가 3-5 사이였던 단어들이 추가된 것을 알 수 있다. 위 15개 단어는 분야를 막론하고 두루 사용되는 표현이라고 추측해볼 수 있다.

(3) N그램 추출(두 단어 / 세 단어) - AntConc

(1)과 (2)에서 공통적으로 많이 등장한 ‘수’, ‘있다’나 ‘본’, ‘연구’ 등의 상위 빈도 공기어들은 실제로 문장 속에서도 많은 경우에 함께 등장하는 것으로 관측되었다. 이를 좀 더 명확하게 알아보기 위하여 AntConc의 N그램 추출 기능을 사용하였다.

인문학	n=2	n=3	사회과학	n=2	n=3	어문학	n=2	n=3
3434 수 있다	646 할 수 있다	4463 수 있다	791 할 수 있다	3673 수 있다	734 할 수 있다			
2136 수 있는	468 볼 수 있다	2714 수 있는	674 볼 수 있다	1402 수 있는	567 볼 수 있다			
1160 할 수	418 알 수 있다	1348 할 수	399 수 있을 것	1169 할 수	536 알 수 있다			
811 볼 수	349 수 있을 것	1107 것으로 나	332 알 수 있다	938 볼 수	248 수 있을 것이다			
636 알 수	162 할 수 있는	945 볼 수	281 영향을 미	900 알 수	188 것을 알 수			
573 수 있을	149 년 월 일	860 년 월	251 년 월 일	407 수 없는	167 있음을 알 수			
530 년 월	93 들 수 있다	741 영향을 미	189 할 수 있는	401 년 월	153 년 월 일			
525 것이 아니	92 것을 알 수	718 본 연구는	153 영향을 미	382 에 대한	147 확인할 수 있다			
467 수 없는	87 확인할 수	678 수 있을	152 미치는 것	379 수 있을	123 상 보조 용언			

자연과학	n=2	n=3	공학	n=2	n=3	의약학	n=2	n=3
2041 수 있다	289 알 수 있다	1437 수 있다	229 알 수 있다	1419 수 있다	195 할 수 있다			
919 수 있는	256 볼 수 있다	738 수 있는	169 할 수 있다	1204 수 있는	149 볼 수 있다			
462 알 수	232 할 수 있다	381 알 수	139 볼 수 있다	852 본 연구는	137 수 있을 것			
442 할 수	192 C U E	373 본 연구에	94 확인할 수	695 본 연구에	131 통계적으로			
406 본 연구에	192 U C U	367 할 수	93 것을 알 수	674 것으로 나	120 것으로 알			
364 볼 수	153 확인할 수	239 볼 수	67 것을 확인	642 본 연구의	117 영향을 미			
339 수 있도록	144 E U C	234 수 있었다	65 확인할 수	607 본 연구에	116 유의한 차			
306 확인할 수	144 U E U	233 과 같이	64 수 있다 또	486 영향을 미	115 알 수 있다			
302 예를 들어	129 β β β	217 확인할 수	63 알 수 있었	481 유의한 차	114 확인할 수			
295 수 있었다	123 것을 알 수	202 수 있도록	61 그림 와 같	467 수술 후	110 일반적 특			

그림 5 : N그램(n=2, n=3) 추출 상위 10개 표현

[그림 5]에서 볼 수 있듯이, 계열에 상관없이 ‘-수 있다’의 형태가 가장 빈번하게 등장하였다. 또한 ‘본 연구에서는’이나 ‘본 연구는’과 같이 논문에서 보편적으로 쓰이는 표현이 고빈도로 나타났다. 다만 자연과학 분야의 경우 전처리 과정에서 특수문자로 처리되지 못한 수학기호들이 상위 키워드로 선정되는 문제가 발생하였다. 이 점을 제외하고 살펴볼 때, 공학 분야에서는 다른 분야와 달리 ‘과 같이’와 ‘그림 와 같이’가 상위 10개의 n-gram에 포함되었다는 점이 특징적이다. 이를 통해 공학 분야에서 표나 그림을 인용하여 설명하는 부분이 상대적으로 많다는 점을 추측할 수 있다. 또한 각각 문과, 이과 코퍼스에 속하는 사회과학과 의약학에서 ‘영향을 미치는’이 고빈도로 나타났다는데, 두 분야의 논문 작성 방식에 어떤 공통점이 있는지 살펴볼 필요가 있다. 또한 의약학에서 다른 계열과 다르게 ‘통계적으로 유의한 차이가’가 고빈도 표현으로 집계되었는데 이는 의약학

에서 통계의 합리성을 검증하는 논리 전개 방식이 사용됨을 시사한다.

(4) 한계점

6개 계열에 대하여 포괄적인 조사를 마친 결과 생각보다 눈에 띄는 차이점이 나타나지 않았다는 문제 상황에 직면하였다. 따라서 추후 연구에서는 두 가지 접근 방식을 통하여 조금 더 세밀하게 결과를 도출하고자 하였다.

첫째는 문/이과 통합 코퍼스를 구축하는 것이다. 이미 코퍼스를 구축하는 과정에서 최대한 많은 학과를 확보하여 코퍼스의 객관성을 확보하였고, 각 계열별 논문의 개수를 동일하게 일치시켰다는 점에서 특정 학문에 편중되는 현상을 어느 정도 방지했다. 이러한 요소들이 각 논문들의 대표성을 어느 정도 보장해준다는 가정 하에 총 900개 논문들로 구성된 문/이과 코퍼스 구축이 가능해진다. 이를 분석하면 학문의 융합과 통섭이 중요한 화두로 제기되는 시기에 과연 학문간 차이가 실제로 좁혀지고 있는지, 아니면 여전히 문/이과의 성향 차이라는 것이 존재하는지를 알아볼 수 있을 것이다.

둘째는 계열을 총 4개의 그룹으로 나누어 다시 한 번 세부적으로 분석하는 방법이다. 이 과정에서 본 연구팀은 인문과학·어문학, 사회과학, 의약학, 자연과학·공학 총 4개 그룹으로 코퍼스를 나누었다. 구분한 기준은 학문적 성격의 유사성 혹은 차이점이다. 사회과학과 의약학은 심리학과 신경과학이라는 유사한 학문을 각기 포함하고 있다. 또한 방법론적으로 사회과학 분야에서 수치 정보나 통계를 많이 사용한다는 점은 널리 알려져 있다. 이것이 사회과학의 이과적 측면을 반영하고 있다고 판단하였다. 반면 의약학 관련 논문 중에서는 임상 실험이나 환자의 상태에 관한 논문이 다수 존재하였다. 이러한 특징이 인문(人文)적 성격을 강하게 반영하고 있다고 판단하였다. 물론 이때 코퍼스 개수가 달라진다는 문제가 발생하는데, 이는 최초 wordlist를 분석할 때와 마찬가지로 상대적 비율을 함께 제시하여 해결할 수 있을 것이라 생각한다.

2. 문/이과 통합 분석

(1) 문/이과 wordlist 추출 비교

문과와 이과의 논문 텍스트에서 다음과 같은 유의미한 공통점 및 차이점을 찾을 수 있었다. 먼저 문과와 이과의 상위 200개의 고빈도 단어를 추출하여 비교해본 결과, 상위 고빈도 단어들 중에는 공통적으로 나타나는 단어들이 많았다. 이는 문과와 이과에서 쓰이는 표현이 전체적으로 봤을 때 큰 차이를 보이지 않음을 의미한다. 상위 고빈도로 집계된 단어들은 문·이과에 상관없이 논문 텍스트에서 자주 쓰이는 표현들이다. 예를 들어 문·이과 모두에서 1, 2위에 집계된 '수'와 '있다'의 경우, 앞에 용언과 다양하게 활용되어 'ㄹ 수 있다'라는 표현으로 매우 빈번하게 쓰인다. 하지만 이때 고려해야 할 점은 같은 개수의 논문을 수집하였으나 결과적으로 문과가 이과의 약 2배에 해당하는 어절 수의 코퍼스가 구축되었다는 것이다. 이는 문과가 이과에 비해 논문의 길이가 대체로 길기 때문이다. 이과는 수식, 그림 등을 사용하여 실험이나 관찰의 과정 및 결과를 간결하게 기술하는 반면, 문과는 연구한 바에 대해 장황한 설명을 덧붙이는 경우가 많다. 코퍼스를 구축할 때 불필요한 정보들을 제거하는 과정에서도 이과가 문과보다 인용이나 각주, 참고문헌 등이 훨씬 적으며, 그 작성 양식이 간단함을 확인할 수 있었다. 코퍼스 구축 후 문과와 이과 간의 어절 수를 동일하게 맞추어 분석하는 방법도 고려하였으나, 이것도 문·이과 논문 텍스트의 분명한 차이점이라고 생각하여 수집한 코퍼스를 그대로 사용하였다. 하지만 절대 빈도만 보고 분석할 경우 문·이과 간의 차이가 어절 수에 의해 비롯되기 때문에 상대 빈도를 살펴볼 필요가 있다고 판단

했다. 따라서 고빈도로 나타나는 단어들에 대해 문과와 이과 간의 상대 비율을 구하여 구체적으로 빈도 차이가 있는지 알아보고자 한다.

문이과 공통					3390023.00000	1662741.00000	상대 비율	상대비율
word	rank(문)	freq(문)	rank(이)	freq(이)	문과 비율	이과 비율	문/이	이/문
수	1	33529	1	15537	0.00989	0.00934	1.05846	0.94477
있다	2	26212	2	13572	0.00773	0.00816	0.94728	1.05565
대한	3	21823	4	10720	0.00644	0.00645	0.99849	1.00152
있는	4	17700	6	7648	0.00522	0.00460	1.13513	0.88095
이	5	14142	10	5402	0.00417	0.00325	1.28404	0.77879
그	6	13968	26	3336	0.00412	0.00201	2.05367	0.48693
의	7	11672	5	9638	0.00344	0.00580	0.59399	1.68353
것이다	8	10936	43	2587	0.00323	0.00156	2.07340	0.48230
및	9	9844	3	11162	0.00290	0.00671	0.43257	2.31179
이러한	10	9342	33	3004	0.00276	0.00181	1.52532	0.65560
것으로	11	9192	8	6205	0.00271	0.00373	0.72659	1.37629

그림 7 : 문/이과 코퍼스 wordlist 추출 결과

먼저 눈에 띄는 차이는 '및'의 사용에 대한 것이다. '및'은 문장에서 두 말을 이어 주거나 열거해 주는 말이다. '및'은 이과에서 총 11162번, 문과에서 9844번 집계되어 이과가 문과보다 약 2.3배 사용됨을 알 수 있다. 이는 문과에서는 단어를 나열할 때 '및'뿐만 아니라 '그리고'나 '또'와 같은 다른 단어도 다양하게 사용함을 보여준다. 또한 '및'은 셋 이상보다 두 말을 이어줄 때 사용하는 경우가 많은데, 이과에서는 '활용 및 현황', '예측 및 분석'과 같이 두 단어를 연결하는 경우가 많음을 알 수 있다.

다음으로 문과가 이과에 비해 '것이다'가 2배, '것은'이 3배 이상 사용한 것을 알 수 있었다. 이는 '-는 것이다'라는 구문과 연관성이 있다. 박나리(2012)¹⁾에 따르면, '-는 것이다' 구문은 기존 선행 연구들에서 강조와 단언의 기능을 갖는 것으로 기술되어 왔다. '것'은 의존 명사로서 앞의 관형형 어미에 따라 다양한 문맥 의미를 표현하는데 특히 논문에서 '-는 것'은 단순 진술뿐만 아니라 필자의 주장을 강조하거나 단정적으로 표현할 때 자주 사용된다. 그리고 '-을 것이다'는 미래 시제를 나타내거나 추측, 가능성, 의지를 드러낼 때 주로 쓰인다. 이는 필자의 주관적인 판단이나 태도가 텍스트에 반영되어 있음을 보여준다. 논문이 연구 내용을 정리, 분석하여 자신의 입장을 밝힌 글이라는 점을 고려했을 때, 이러한 표현이 자주 쓰이는 것은 자연스러운 현상이다. 그런데 중요한 점은 왜 이과보다 문과에서 이러한 표지가 빈번하게 쓰이는지에 대한 것이다.

그 이유는 문과의 논리 전개 방식과 연관 지을 수 있다. 최도식(2015)²⁾에 따르면, 인문 계열의 글쓰기는 인간의 가치탐구와 표현활동을 대상으로 하여, 인간 삶의 본질적인 문제를 다룬다. 그래서 글의 논리 전개 방식이 대체로 사유의 논리성, 즉 인간 혹은 삶의 문제에 대해 능동적으로 생각하고 해결하기 위한 논거를 제시하는 과정과 관련되어 있다. 따라서 개인의 가치판단이 개입되며, 주장 혹은 결론을 뒷받침할 수 있는 논증을 펼치는 것이 중요하다. 한편, 사회 계열 글쓰기의 경우 다양한 사회 현상들을 다루는데, 최도식(2015)³⁾에서는 사회 현상에 접근하는 방식이 크게 세 가지로 나누어진다고 보았다. 사회 현상의 사실 관계나 인과 관계를 탐구하는 방법, 바람직한 사회 상태가 무엇인가에 대해 규명하는 방법, 이상적 목표를 전제하고 이를 위한 방안을 제시하는 방법으로 설명한다. 이 과정에서 사회계열은 통계와 같이 객관적인 자료를 제시하는 경우가 많기 때문에 자료의 적합성과 합리성을 입증할 필요가 있다. 최도식(2015)은 학술 텍스트를 대상으로 글쓰기의 특징을 설명한 것이 아니지만, 본 연구에서 나눈 어문학과 인문과학은 앞서 설명한 인문

1) 박나리, 「-는 것이다 구문 연구」, 국어학회, 2012, p225

2) 최도식, 「계열별 글쓰기의 특성과 교육방법적 모형 개발」, 사학과언어학회, p186-187

3) 최도식, 「계열별 글쓰기의 특성과 교육방법적 모형 개발」, 사학과언어학회, p187

계열 글쓰기와, 사회과학은 사회 계열의 글쓰기와 맥락은 같이한다고 볼 수 있다. 인문계열과 사회계열은 공통적으로 필자의 주관적인 생각이 글쓰기에 중점적으로 나타난다는 것을 알 수 있는데, 이를 '것이다'와 '것은'의 높은 빈도와 연관 지을 수 있을 것이다.

이공계열의 글쓰기 전개 방식과 비교하면 더 이해가 쉬울 것으로 판단된다. 최도식(2015)⁴⁾에서는 이공계열의 글쓰기는 과학적 방법, 실험적 증명을 기반으로 하여 어떤 현상을 관찰한 결과나 실험 결과를 주로 다룬다고 설명한다. 그래서 이공계열에서는 연구 과정 및 결과를 설명하고, 그 결과가 도출된 원인을 분석하는 경우가 많다. 이때 어떤 현상의 인과 관계에 대해 논리적으로 그럴듯하고 타당한 설명을 생각해내는 논리적 추론, 현실 세계에서 적용되는지 맞추어보는 경험적 검증 과정이 포함된다. 그런데 이과 계열의 경우, 독자들의 입장에서 결과 자체는 물론, 결과에 대한 해석 또한 과학적·합리적으로 이해되어야 한다. 따라서 문과와 비교했을 때, 필자의 주관성이 드러나는 표현이 상대적으로 덜 사용된다. 게다가 문과는 상대 빈도수가 높음은 물론이고 '-는 것'의 구문이 다양한 용언과 함께 쓰임을 확인할 수 있었으나, 이과는 '-는 것이다'라는 구문은 잘 사용되지 않았다. '-을 것이다'는 결과에 대한 추측과 미래형 표현으로서 사용됨을 볼 수 있었다.

antconc의 고빈도 단어 추출 결과, 이과에서 '의', '를', '에', '로', '와', '은', '과', '을' 등의 조사상위 빈도에 집계된 것을 확인할 수 있다. 이는 영문과 숫자를 지웠기 때문에 발생한 현상으로, 이과가 문과에 비해 영문과 수식을 많이 사용함을 알 수 있다. 또한 이과계열 논문에서는 독자의 이해를 돕기 위해 그림이나 표를 통해 결과를 도식화하는 경우가 대부분이었다. 이때 한국어 논문 이더라도 그림과 표를 figure(혹은 fig)이라고 쓰는 경우가 많았는데, 이들을 논문상에서 언급할 때는 당연히 조사가 함께 쓰인다. 그런데 figure 또한 전처리 과정에서 삭제되었으므로 조사만 따로 집계된 현상이 문과보다 더 많아진 것으로 생각된다.

이어서 유의미하게 살펴볼 차이점은 부정어에 대한 것이다. 상위 고빈도 단어 200개 의 결과에서 '아니라'는 문과가 약 1.8배, '안'이 1.6배 많이 출현했으며, '없다', '않는', '않고'는 문과에서만 나타났다. 이를 통해 문과가 이과보다 부정 표현의 사용이 두드러짐을 알 수 있다. 물론 더 아래 빈도를 살펴보면 이 단어들도 이과에 존재하겠지만, 문과에서 상대적으로 더 많이 사용된다는 것은 분명한 사실이다. 이는 문·이과 논문의 논리 입증 방식이 다르기 때문으로 볼 수 있다. 이과가 수치나 실험적 결과로서 객관성을 입증하는 반면, 문과에서는 타당한 논거를 들어 논리적으로 입증하기 때문에 기존 연구의 논리적 결점을 찾는 과정에서 부정 표현을 많이 사용하는 것이다. 이때의 부정 표현은 자신의 주장을 강화하기 위한 근거가 되기도 한다, 그러므로 주장의 설득력을 얻기 위한 전략상의 차이에서 이러한 부정어 사용 빈도 차이가 발생한다. 한편, 구어나 일반 텍스트에서는 부정 표현이 돌려말하는 등의 다양한 의미로 사용될 수도 있다. 하지만 논문 텍스트는 필자의 연구 내용과 견해를 명확히 밝힐 필요가 있기 때문에 그러한 부정어 사용은 거의 관찰되지 않는다.

(2) 문/이과 통합 코퍼스 품사별 추출 비교 - mecab

부사

문·이과 간의 또 다른 차이점은 바로 '부사'와 관련한 것이다. 문과가 이과에 비해 접속 부사, 일반 부사가 훨씬 다양하고 빈번하게 사용되었음을 확인하였다.

4) 최도식, 「계열별 글쓰기의 특성과 교육방법적 모형 개발」, 사학과언어학회, p188

문/이과 접속부사 공통					3390023	1662741	상대비율	상대비율
word	rank(문)	freq(문)	rank(이)	freq(이)	문과 비율	이과 비율	이/문	문/이
및	1	10642	1	12185	0.00314	0.00733	2.33443	0.42837
그리고	2	6560	3	1990	0.00194	0.00120	0.61848	1.61686
또한	3	6385	2	3932	0.00188	0.00236	1.25554	0.79647
즉	4	5568	7	1394	0.00164	0.00084	0.51044	1.95911
그러나	5	5209	5	1534	0.00154	0.00092	0.60041	1.66552
따라서	6	3444	4	1817	0.00102	0.00109	1.07565	0.92967
또는	7	3024	6	1456	0.00089	0.00088	0.98165	1.01869
혹은	8	2385	8	763	0.00070	0.00046	0.65225	1.53315
역시	9	2254	10	486	0.00066	0.00029	0.43960	2.27478
그런데	10	1400	12	151	0.00041	0.00009	0.21990	4.54750
하지만	11	1384	9	538	0.00041	0.00032	0.79255	1.26176
오히려	12	1118	13	150	0.00033	0.00009	0.27354	3.65572
그러므로	13	944	11	270	0.00028	0.00016	0.58314	1.71487

그림 8 : 문/이과 코퍼스 접속부사 빈도 분포

먼저 접속 부사는 글의 흐름을 나타내는 대표적인 표지이다. 접속 부사만 따로 분류하여 그 양상은 살펴본 결과, 대부분의 접속 부사가 문과에서 훨씬 많이 쓰였음을 알 수 있었다. 아래 표에서 '및', '또한', '따라서', '또는', '하지만'을 제외한 모든 부사들이 문과가 이과에 비해 1.5배 이상 많이 사용되었다. 이과에서 대표적으로 쓰이는 부사는 나열 시에 사용되는 '및', '또한', '또는', '그리고', 결론을 제시하는 '따라서', 역접이나 대조의 '하지만', '그러나', 재진술의 '즉'이 있다. 이과의 접속 부사 빈도에서 7위의 '즉'은 빈도가 1394번인데, 8위의 '혹은'은 763번인 것으로 나타나는데, 이는 이과에서는 자주 쓰이는 특정 접속 부사를 제외하면 다양한 접속부사를 사용하지는 않는 것으로 해석할 수 있다. 한편 이과의 고빈도 접속부사는 문과도 마찬가지로 나타나지만 강조의 '역시', '이른바', '더욱이', 역접(대조)의 '그런데', '오히려', '그렇지만', 인과의 '그러므로', '그래서', '왜냐하면', '그리하여', 이외에도 '다만', '요컨대', '소위' 등 훨씬 다양한 부사가 사용되었다. 따라서 문과는 문장과 단락들을 응집시켜 하나의 통일성 있는 내용을 전개하는 경향이 강함을 알 수 있다. 이는 문과에서 '이렇게', '이러한', '이런', '그러한', '그것은', '앞서', '그는', '그의', '이것은', '이처럼' 등의 지시어를 쉽게 찾을 수 있다는 것과도 일맥상통한다.

문/이과 일반부사 공통					3390023	1662741	상대비율	상대비율	문과만		
word	rank(문)	freq(문)	rank(이)	freq(이)	문과 비율	이과 비율	이/문	문/이	word	rank(문)	freq(문)
다	1	19798	1	5483	0.00584	0.00330	0.56464	1.77103	그렇게	60	458
더	2	6735	3	2812	0.00199	0.00169	0.85125	1.17475	결코	67	399
가장	3	4754	2	3127	0.00140	0.00188	1.34106	0.74568	과연	68	395
같이	4	4232	4	2397	0.00125	0.00144	1.15478	0.86596	예컨대	71	374
특히	5	4091	6	1539	0.00121	0.00093	0.76699	1.30381	분명히	74	359
함께	6	3061	10	1189	0.00090	0.00072	0.79195	1.26271	마치	75	349
모두	7	3058	5	1661	0.00090	0.00100	1.10741	0.90300	심지어	77	317
또	8	2987	18	568	0.00088	0.00034	0.38770	2.57934	적어도	79	305
보다	9	2588	8	1503	0.00076	0.00090	1.18406	0.84455	끊임없이	82	295

그림 9 : 문/이과 코퍼스 일반부사 빈도 분포

일반 부사의 경우도 비슷하다. 먼저 '다'가 양쪽 코퍼스 모두에서 최고 빈도로 집계된 것은 코퍼스 상의 한계이다. pdf 파일을 텍스트 파일로 변환하는 과정에서 의도치 않게 종결어미에서 '다'가 띄어쓰기가 되는 현상이 많이 발생했다. 이를 최대한 수정하고자 노력했으나 미처 발견하지 못한 '다'가 남아 일반 부사로 집계되었다. 이를 제외하고 봤을 때, 상위 10위 이내로 출현한 일반 부사들은 문과와 이과의 사용빈도가 비슷했다. '더', '가장', '같이', '특히', '함께', '모두', '보다', '매우', '잘' 등이 그 예이다. 강조할 때 사용하는 표현은 '가장', '매우', '더욱'의 순으로 자주 사용된다는 공통점이 있었다.

일반 부사가 접속 부사보다 단어 양상이 훨씬 다양하다보니 문·이과의 차이점이 접속부사에 비해서는 명확하게 드러나지 않았다. 그래도 문과가 이과에 비해 다양한 일반 부사를 사용한다는 점은 분명하다. '또', '다시', '어떻게', '바로', '이미', '달리', '안', '물론', '우선', '실제로', '여전히', '그대로', '곧', '없이', '더욱' 등이 문과가 이과보다 1.5배 이상 쓰인 부사들이다. 특히 '바로'는 1700번 출현해 문과 일반부사에서 14위의 빈도를 나타냈는데, 이과는 205회 출현하여 40위 빈도를 나타냈다. 즉, '바로'가 문과에서 4배가량 많이 쓰인 것인데, 이는 문과에서 '바로 -이다', '바로 -이라는 것이다'라는 구문으로 강조 시에 자주 등장하기 때문이다. 4배 이상 많이 집계된 '곧', '분명', '사실'도 마찬가지로 맥락이다.

또한 이과에만 나타나는 부사는 거의 없었으나, 문과에만 나타나는 일반 부사는 '그렇게', '결코', '과연', '예컨대', '분명히' 등 다수 존재했다. 그런데 이와 대조적으로 이과가 문과보다 두드러지게 많이 쓰였다고 볼 수 있는 부사는 '각각'밖에 없었다. '각각'은 이과에서 실험 등에서 세부 조건을 나눌 때 자주 사용하는 표현으로, 이과 계열의 특징을 반영하는 것이다. 결론적으로, 문과는 이과에 비해 더 다양한 부사를 많이 사용하여 내용을 강조하거나, 글의 흐름을 매끄럽게 전개하고자 함을 알 수 있다.

동사/형용사

품사 중 동사와 형용사를 분류하여 빈도와 상대 비율을 확인해보았다. 본 연구에서는 용언 중에서 형용사보다 동사를 중점적으로 살펴보기로 했는데, 이는 연구 내용을 설명하고 필자의 견해를 밝히는 논문의 특성상 동사의 비중이 높기 때문이다. 상태나 성질을 나타내는 형용사는 '있다', '없다', '높다', '낮다', '많은', '큰', '다양한', '다른', '새로운', '모든', '유의한', '같다', '중요한' 정도를 제외하면 높은 빈도의 형용사가 나타나지 않았다. 이 단어들은 대부분 문과와 이과에서 상대 빈도가 비슷하게 나타나는데, 이는 형용사의 사용에 있어서 문·이과 간 차이점이 뚜렷하지는 않다는 점을 보여준다.

그래도 차이점으로 보이는 것은 '없다'가 문과에서만 고빈도로 나타난다는 점, '유의한'이 이과에서만 고빈도로 나타난다는 점이다. '유의하다'는 의미나 뜻이 있다는 사전적 의미를 갖는데, 실험의 결과나 수치상 자료가 유의미하다는 맥락에서 사용된다. 사회 과학 계열도 연구에서 통계적 방법론을 이용하기 때문에 '유의하다'라는 표현을 많이 사용한다. 하지만 문과 전체 코퍼스 분석한 것이기 때문에 문과에서 '유의한'은 빈도 순위가 낮게 나온 것으로 생각된다. 반면 이과 전반은 물론, 특히 의학에서 통계적 수치 자료의 유의성을 판단하는 연구 방법을 주로 사용하기 때문에 위의 결과가 도출된 것으로 보인다.

동사의 경우 활용형을 고려하여 어간 추출을 통해 빈도를 살펴보았다. 이 결과에서 아쉬운 점은 '어떤 사실이나 현상이 존재하는 상태'를 의미하는 형용사로서의 '있다'까지 포함되어 '있-'이 문이과 모두에서 최고 빈도로 나타났다는 것이다. 물론 동사, 보조동사로서의 '있다'도 '수'와 함께 매우 자주 쓰이기 때문에 고빈도에 포함되었지만, 형용사를 제외하지 못했다는 점이 한계로 남는다. 한편, 상위 10위 이내로 집계된 단어들, '하-', '되-', '보-', '받-', '미치-', '가지-', '보이-', '갖-', '살펴보-'는 문이과 상대 비율이 비슷했다. 그 이후에 나타난 동사들 중 차이점을 위주로 살펴보면, 먼저 '들-'은 문과에서 조금 더 많이 사용되었다. 이는 문과에서 예시를 드는 경우가 더 빈번하기 때문이다. 그리고 '읽-'이 문과에서 약 4.5배가 더 높은 빈도가 보인 것은 '글이나 언어를 읽다'라는 맥락이 자주 등장하기 때문이다. 연구 대상이나 방법에 있어서 문과는 문학 작품, 언어 자료, 사회 현상의 분석 자료 등의 텍스트를 다루지만, 이과는 자연 현상, 기계나 장치 등을 다루기 때문에 이러한 차이가 나타난다. '쓰-'가 문과에서 3배 이상 나타난 것도 같은 해석을 적용할 수 있다.

한편, 이과에서 '나타내-'가 문과에 비해 2.5배 이상 사용되었다. 이는 실험 및 관찰 결과를 설명할 때 '나타낸다', '나타내었다', '나타났다', '나타낸 것으로 보인다' 등의 표현이 빈번하게 쓰이기 때문이다. 같은 맥락에서 '얻-', '밝히-'도 이과가 더 많이 사용되었다. 한편, 어떤 정도나 범위에 미쳤다는 표현으로서 '이르-', 무엇을 무엇으로 여기거나 가정하는 표현으로서 '삼-' 등이 문과에서 더 많이 사용된다. 그리고 문과는 개인이나 사회 문제를 다루기 때문에 '살-', '느끼-', '겪-', '먹-', '사-', '듣-', '맺-', '죽-', '믿-', '만나-' 등과 같이 사고와 행동에 관련한 동사들은 문과에서 주로 나타났다. 이과의 경우 '풀-', '낮추-', '가하-', '높이-' 등과 같이 실험 및 관찰의 조건에 관련된 표현이 더 많이 나타났다.

명사

마지막으로 살펴볼 품사는 명사이다. 명사의 경우, 연구 주제에 따라 다양하게 나타나지만 논문에서 자주 사용되는 명사들도 있다.

문과·이과 공통 명사					3390023	1662741	상대비율	상대비율	문과만			이과만		
word	rank(문)	freq(문)	rank(이)	freq(이)	문과 비율	이과 비율	이/문	문/이	word	rank(문)	freq(문)	word	rank(이)	freq(이)
연구	1	26512	1	20071	0.00782	0.01207	1.54349	0.64788	말	12	8313	환자	6	6779
사회	2	15565	61	2932	0.00459	0.00176	0.38405	2.60380	문화	13	8179	수학	7	6557
경우	3	10583	5	7580	0.00312	0.00456	1.46029	0.68480	도서관	14	8019	치료	36	3615
분석	4	10012	4	8168	0.00295	0.00491	1.66331	0.60121	국가	21	7239	간호	38	3541
사용	5	9839	3	8738	0.00290	0.00526	1.81067	0.55228	사람	27	6044	값	44	3362
의미	6	9753	86	2389	0.00288	0.00144	0.49941	2.00237	역사	30	5862	간호사	56	2987
관계	7	9662	30	3850	0.00285	0.00232	0.81240	1.23092	인간	34	5149	군	57	2985
문제	8	9538	11	5937	0.00281	0.00357	1.26908	0.78797	경제	37	5045	대상자	58	2962
결과	9	9294	2	10008	0.00274	0.00602	2.19545	0.45549	존재	39	4996	유의	62	2905
학습	10	8787	35	3628	0.00259	0.00218	0.84179	1.18794	영어	40	4973	감소	63	2834

그림 10 : 문/이과 코퍼스 명사 빈도 분포

먼저 문·이과에서 공통적으로 '연구'가 최고 빈도로 나타났는데, 이는 논문에서 '본 연구(에서)'와 같은 표지를 많이 쓰기 때문이다. 빈도 측면에서 이과가 '연구'라는 단어를 문과보다 약 1.5배 많이 사용했는데, 이는 '본'이 이과에서 1.6배 정도 많이 사용된 것과도 일치한다. '사회', '의미', '지역', '자신'은 문과에서 많이 사용되는 경향이 강했으며, 이과에서 '분석', '사용', '결과', '평가', '차이', '방법'의 빈도가 높게 집계된 것은 이과의 연구 방법과 연결지을 수 있다. 사실 이 정도를 제외하고는 문·이과 명사 사용 양상은 비슷하게 나타났다. 그 중 예상과 달랐던 것은 '학습'과 '교육'의 빈도가 이과에서 생각보다 높았다는 점이다. 사회 계열 내에 교육학이 존재하지만, 교육학은 전공별로 세분화되어 있기 때문에 이런 결과가 나타난 것으로 보인다. 즉, 이과 계열의 논문에 교육학 관련 논문들이 다수 포함되었다는 것이다. 이는 논문의 주제별 분류가 명확하지 않을 수 있음을 보여주기도 한다. 그런데 '학습'의 경우, 교육학에서의 '학습' 개념뿐만 아니라 기계, 생산 및 작업의 학습 (효과) 등과 관련해서 공학 계열에서, 의료 전문직이나 환자의 행동 및 운동 학습과 관련해서 의학 계열에서도 자주 쓰이기도 한다.

다음으로 문과 혹은 이과에서만 상대적으로 높은 빈도로 나타나는 명사의 경우, 연구 주제와 관련성이 높았다. 예를 들어, 문과에서는 '말', '문화', '도서관', '국가', '사람', '역사', '인간', '경제', '존재', '영어', '정책', '언어', '세계', '이론', '주장', '정치', '지원', '논의', '공간', '정부' 등이 사용되었다. 이과에서는 '환자', '수학', '치료', '간호', '값', '간호사', '군', '대상자', '유의', '감소', '건강', '수술', '세포', '운동', '의료', '검사', '점수', '통계', '데이터', '시스템' 등이 사용되었다. 이를 통해 문·이과에서 연구 주제로 무엇을 주로 다루는지 파악할 수 있을 것이다.

3. 계열별 분석

(1) 계열별 명사/동사 비교

명사

의학, 자연과학과 공학, 사회과학, 인문과학과 어문학 이렇게 네 분류로 나누어 계열별 분석을 진행하였다. 의학 계열의 경우 형태소의 개수가 1,403,922개, 자연과학과 공학은 1,713,296개, 사회과학은 2,503,263개, 인문과학과 어문학은 4,215,583개가 추출되었다. 전체에서 의학은 14%, 자연과학과 공학은 17%, 사회과학은 25%, 인문과학과 어문학은 43%를 차지하고 있다. 이는 형태소의 개수이므로 앞의 표에서 제시된 어절 수와는 다르다. 그런데 의학 계열의 경우, 자연과학과 공학을 합친 코퍼스와 비교했을 때 어절 수와 형태소 수에 있어서 큰 차이를 보이지 않는다. 이는 의학 계열이 자연과학, 공학에 비해서 논문의 길이가 길다는 점을 의미한다. 또한 의학 계열은 자연과학, 공학에서 숫자나 수식을 많이 사용하는 것과 달리, 설문조사 등을 통한 통계적 방법론을 다수 이용하는 것으로 보인다. 이러한 이유들로 인해 이과 코퍼스 내에서 의학 계열이 거의 절반에 가까운 비중을 차지하게 되었다. 따라서 이와 같이 계열별 코퍼스 크기의 차이가 존재하기 때문에 문·이과 분석에서와 마찬가지로, 절대 빈도뿐만 아니라 상대빈도를 고려할 필요가 있다.

의학	형태소 개수 : 1403922	자연_공학	형태소 개수 : 1713296	사회 과학	형태소 개수 : 2503263	어문_인문	4215583
kor tag num		kor tag num		kor tag num		kor tag num	
0 연구 NNG 13131		0 연구 NNG 6940		0 연구 NNG 13804		0 연구 NNG 12708	
1 환자 NNG 6548		1 수학 NNG 6556		1 사회 NNG 8513		1 도서관 NNG 8006	
2 결과 NNG 5244		2 사용 NNG 5139		2 관계 NNG 5760		2 의미 NNG 7496	
3 후 NNG 4341		3 분석 NNG 5016		3 결과 NNG 5328		3 사회 NNG 7052	
4 사용 NNG 3599		4 결과 NNG 4764		4 분석 NNG 5299		4 말 NNG 6761	
5 간호 NNG 3540		5 문제 NNG 4754		5 효과 NNG 4864		5 경우 NNG 6193	
6 경우 NNG 3516		6 학생 NNG 4555		6 영향 NNG 4740		6 사용 NNG 6065	
7 점 NNG 3376		7 경우 NNG 4064		7 국가 NNG 4718		7 문화 NNG 5608	
8 치료 NNG 3334		8 방법 NNG 4060		8 경우 NNG 4390		8 점 NNG 5461	
9 분석 NNG 3152		9 과정 NNG 3838		9 문제 NNG 4282		9 문제 NNG 5256	
10 효과 NNG 3109		10 이용 NNG 3641		10 학습 NNG 3815		10 역사 NNG 5155	
11 간호사 NNG 2983		11 학습 NNG 3064		11 사용 NNG 3774		11 학습 NNG 4972	

그림 11 : 계열별 명사 분포

계열별 명사 분석에서 어떤 차이점이 나타나는지 알아보고자 한다. 본격적인 분석에 앞서 짚고 넘어가야 할 것은 형태소 단위로 집계된 결과이기 때문에 ‘하다’ 접사가 붙어 용언으로 쓰이는 경우에 그 어근에 해당하는 명사가 대부분 포함되었다는 것이다. 이는 사실상 동사나 형용사이기 때문에 제외하는 것이 적절하다. 하지만 ‘하다’ 파생어가 다양하게 많이 쓰여 일일이 제외하기가 어려우며, 용언으로서 자주 쓰이는 표현도 유의미하다는 판단 아래 그대로 분석을 진행하였다. 일단 공통적으로 모든 계열에서 최고 빈도로 ‘연구’가 등장한 것을 확인할 수 있다. 이는 논문이 특정 주제의 ‘연구’를 다루기 때문에 당연한 결과이다. 그런데 이과 내에서 자연과학&공학보다도 의학에서 압도적으로 ‘연구’라는 명사가 많이 사용된 것은 주목할 만하다. 자연과학&공학도 상대적으로 문과보다는 ‘연구’가 많이 출현했으나, 의학 계열에서는 코퍼스 크기가 훨씬 큰 사회과학이나 인문과학&어문학보다도 더 많은 빈도가 집계되었다.

이어서 의학 분야에서 나타나는 특징을 기술하고자 한다. ‘환자’, ‘후’, ‘간호’, ‘치료’, ‘간호사’, ‘대상(자)’, ‘군’, ‘건강’, ‘수술’, ‘문항’, ‘운동’, ‘프로그램’, ‘증상’, ‘의료’, ‘병원’, ‘임상’ 등은 의학 분야에서만 나타난 명사들이다. 이것들은 의학 계열에서 다루는 연구 내용과 밀접한 관련이 있다. 의학 계열은 다른 계열과 달리 계열의 특성을 구체적으로 드러내는 명사들이 상위 고빈도 중 다수를 차

지한다. 앞의 단어들 중 '후'가 눈에 띄는데, 이는 의약 계열에서 치료 혹은 임상 실험의 전과 후를 비교하는 경우가 많기 때문으로 추측된다. 그리고 의약 계열에서 설문조사 등의 통계적 방법론을 많이 사용한다는 점을 보여주는 명사들도 쉽게 찾을 수 있다. '효과', '유의', '차이', '요인', '문항', '이상', '점수' 등이 그 예이다. '통계적으로 유의하다'는 표현이 의약학의 ngram에서 자주 등장한 것과도 연결지을 수 있다. 이는 독립 요인, 종속 요인에 따른 두 집단 간의 차이가 유의미하다고 여겨질 때 일반적으로 사용되는 표현이다.

다음으로 자연공학과 공학의 경우를 살펴보면, 먼저 '수학'이 2위의 고빈도 단어로 출현하였고, 주목할 만한 부분은 교육학에 등장할 법한 명사들, '학생', '교사', '교육'이 고빈도로 출현했다는 점이다. 이는 앞서 문·이과 분석에서도 찾아볼 수 있었다. 교육학은 주제 분류로는 사회 과학에 속하지만, 교육학의 특성상 모든 전공에서 교육 관련 연구를 다룰 수밖에 없다. 이는 논문을 수집하는 과정에서 이과라 할지라도 교육적 측면을 다룬 연구 논문이 포함될 수밖에 없음을 의미한다. 그렇지만 이 명사들이 고빈도로 나타난 것으로 보아 의도치 않게 관련 논문이 다수 포함되었음을 알 수 있다. 이는 본 연구의 한계점으로 지적할 수도 있다. 자연과학과 공학의 특성에 보다 밀접한 논문을 수집하여 코퍼스 분석의 정확성을 높일 수도 있었기 때문이다. 그러나 이것 또한 과정에서 연구자의 자의적인 판단이 지나치게 개입할 수 있다. 이와 같은 문제를 방지하고자 코퍼스크기를 충분히 확보하였으나 부족한 점이 있었다고 판단된다. 또한 이러한 한계는 여러 전공이 복합적으로 결합된 연구의 경우 문과와 이과 혹은 계열별로 분류하기 어렵다는 점을 시사한다.

자연 과학과 공학에서 다른 계열보다 빈번하게 사용된 명사는 '사용', '분석', '방법', '과정', '이용', '평가', '제시', '활용', '확인' 등이다. 이는 이 두 계열에서 기계나 장치 혹은 실험적 방법, 분석(해석) 방법을 다양하게 사용하기 때문이다. 어떤 사물이나 방법을 '사용', '이용', '활용'하여 연구대상을 '분석', '평가'하고, 결과를 '확인', '제시'하는 것은 자연 과학&공학 계열에서 일반적으로 포함되는 과정이다. 이외에도 '그림', '값', '기술', '해결', '모형', '적용', '비교', '데이터', '과제', '실험' 등이 이 두 계열에서 보다 높은 빈도를 보였다.

다음은 사회 과학 계열의 명사 분석이다. '사회', '관계', '영향', '국가', '경제', '정책', '조직', '기업', '협력', '참여', '개인' 정도가 사회 과학 계열에서만 자주 쓰인 명사들이다. 이를 통해 사회 과학이 사회 현상을 연구 대상으로 삼는다는 점, 연구 대상과 관련하여 관계성과 영향력에 초점을 둔다는 점, 어떤 연구 주제를 주로 다루는지 등을 알 수 있다. 사회 과학은 같은 문과라도 인문과학&어문학에 비해 수치적 자료를 많이 사용하며, 변수를 설정하고 집단 간의 비교를 통해 유의미한 결과를 도출하는 실증적 방법론을 이용한다. 이는 '집단', '요인', '변수'의 빈도수가 높게 나타난 것과 관련이 있다. 의약학의 경우에는 비슷한 경향을 보이므로 사회 과학과의 차이가 명확하진 않았으나, 인문과학&어문학이나 자연과학&공학과는 분명한 차이를 보였다.

마지막으로 인문과학&어문학의 경우 2순위로 '도서관'으로 매우 높게 집계되었다는 점이 눈에 띈다. '도서관'이라는 명사는 연구 주제를 직접적으로 보여주기 때문에 고빈도로 출현하리라고는 예상치 못하였다. 그래서 코퍼스를 확인해본 결과, '도서관' 관련 논문이 몇 개 포함되었는데 이 논문들의 길이가 매우 긴 편이었다. 게다가 '도서관'이라는 단어는 동의어나 유의어가 없기 때문에 반복적으로 계속 쓰이는 경향이 강했다. 이러한 점을 미연에 방지하고자 특정 전공에 편중되지 않도록 논문들을 수집하였으나, 그 안에서도 특정 연구 주제로 편중될 수 있음은 고려하지 못하였다. 이러한 점이 아쉬움이자 한계로 남는다. 한편, 인문과학은 인간과 인간의 삶에 관련된 문제를, 어문학은 개인 및 사회에서의 언어 현상을 다루기 때문에 '의미', '말', '문화', '역사', '사람', '인간', '자신', '존재', '인식', '영어', '표현', '세계' 등의 명사가 고빈도로 등장한다.

동사

	A	B	C	D	E	F	G	H	I	J	K
1	의학 동사	형태소 개수 : 1403922		자연_공학 동	형태소 개수 : 1713296		사회 동사	형태소 개수 : 2503263		어문_인문 동	형태소 개수
2	kor tag	num		kor tag	num		kor tag	num		kor tag	num
3	0	있 VV 8520		0	있 VV 12310		0	있 VV 17951		0	있 VV 26867
4	1	하 VV 4862		1	하 VV 5243		1	하 VV 8092		1	하 VV 16014
5	2	되 VV 2239		2	되 VV 2422		2	되 VV 3957		2	되 VV 7931
6	3	받 VV 1786		3	나타내 VV 1322		3	미치 VV 2475		3	보 VV 4491
7	4	보 VV 1500		4	보 VV 1074		4	받 VV 2100		4	받 VV 3257
8	5	미치 VV 1387		5	가지 VV 1052		5	보 VV 1969		5	보이 VV 1960
9	6	보이 VV 633		6	만들 VV 855		6	가지 VV 1186		6	갖 VV 1939
10	7	위하 VV 617		7	받 VV 824		7	살펴보 VV 1106		7	들 VV 1833
11	8	나타내 VV 604		8	얻 VV 821		8	보이 VV 982		8	가지 VV 1816
12	9	이 VV 502		9	위하 VV 807		9	갖 VV 947		9	살펴보 VV 1771
13	10	가지 VV 492		10	다루 VV 756		10	나타나 VV 873		10	만들 VV 1677
14	11	나타나 VV 427		11	살펴보 VV 715		11	위하 VV 824		11	나타나 VV 1660
15	12	얻 VV 419		12	갖 VV 688		12	들 VV 809		12	읽 VV 1507
16	13	주 VV 410		13	들 VV 634		13	이루어지 VV 620		13	찾 VV 1309
17	14	살펴보 VV 388		14	대하 VV 628		14	통하 VV 603		14	가 VV 1197
18	15	대하 VV 325		15	보이 VV 602		15	주 VV 602		15	이루 VV 1167
19	16	이루 VV 315		16	미치 VV 584		16	이루 VV 577		16	나타내 VV 1162
20	17	알아보 VV 304		17	구하 VV 550		17	따르 VV 574		17	쓰 VV 1071
21	18	통하 VV 300		18	찾 VV 526		18	두 VV 567		18	두 VV 1026

그림 12 : 계열별 동사 분포

가장 많이 사용되는 동사 1, 2, 3위는 계열별로 차이가 없이 '있다, 하다, 되다'로 나타났다. 문, 이과 동사 비교 부분에서 언급했듯이 있다가의 빈도가 높은 것은 형용사인 '있다'까지 포함되었기 때문이다. 차이점을 중심으로 각 계열별 특징을 살펴보면 다음과 같다.

먼저 의학계열에서는 '받다, 미치다, 보이다, 일으키다, 앓다'가 상대적으로 자주 사용되었다. '받다'는 '치료를 받다', '투석을 받다', '항암화학요법을 받다'와 같이 치료나 처치를 시행하는 맥락이나 '영향을 받다', '스트레스를 받다'와 같은 맥락에서 사용되었다. '미치다'의 경우 '미치는 영향', '미치는 효과', '영향을 미치는 요인' 등의 구성으로 주로 사용되었으며 신체 상태에 영향을 주는 요인을 분석하는 것을 목적으로 하는 의학계열 논문의 특징을 잘 보여준다. '보이다'는 '증상을 보이다', '차이를 보이다', '상관관계를 보이다'와 같이 환자나 피실험자의 상태를 객관적으로 드러내기 위한 서술어이다. '일으키다'의 경우 '감염을 일으키다', '암을 일으키다', '현상을 일으키다'처럼 쓰여 질병이나 증상이 발현되는 것을 보여주는 대표적인 서술어로 사용되었다. '앓다'는 다른 동사들보다 구체적인 의미를 가지는 동사여서 조금 의외로 느껴지는데 의학계열 논문에서는 피실험자들이 실험을 어떤 자세에서 실시했는지가 중요하기 때문에 '앓은 자세로'와 같은 표현이 자주 등장했으며 '앓았다 일어나기'가 실험항목인 경우도 있었기 때문에 '앓다'가 높은 빈도로 나타난 것으로 보인다.

다음으로 공학, 자연과학 계열에서는 '만들다', '다루다', '구하다', '찾다', '풀다'가 상대적으로 많이 사용되었다. '만들다'는 공학계열에서 '자동차를 만들다', '소프트웨어를 만들다'와 같은 표현이나 자연과학에서 '시약을 만들다'에서 알 수 있듯이 실제로 무언가를 생산해내는 공학이나 자연과학의 특징을 보여준다고 할 수 있다. '다루다'는 '혼합 계산을 다루는 것으로', '본고에서 다루고 있는'에서 처럼 논문에서 주제로 하는 어떤 구체적인 수학개념이나 연구대상을 언급하기 위해서 사용되었는데 연구대상이 구체적이며 확실한 공학, 자연과학 분야의 특징을 보여준다고 할 수 있으며 특히 수학교육 분야에서 교과서에서 어떤 수학개념을 '다룬다'는 맥락에서 많이 사용되었다. '구하다', '찾다', '풀다'는 모두 수학과 관련된 맥락에서 나타난다. '구하다'는 '해를 구하다', '부피를 구하다', '찾다'는 '극솟값을 찾다', '답을 찾다', '풀다'는 '방정식을 풀다'와 같이 사용되어 수학적 계산을 요구하는 공학, 자연과학 계열의 특징을 잘 보여준다.

사회계열의 특징적인 동사는 '미치다', '위하다', '통하다', '대하다', '찾다'이다. '미치다'는 주로 '영향을 미치다' 꼴로 사용되며 두 변인간의 상관관계를 논하는 사회과학 논문의 특징을 보여준다. '위하다, 대하다, 통하다'는 '위한, 위하여, 대한, 대하여, 통하여, 통한'의 형태로 주로 사용되며 '위하다'의 경우 사회현상의 원인과 결과를 분석하기 위해 '~을 위하여'와 같이 사용되는 경우가 많고 '대하다'는 '북한에 대한', '범지구적인 문제에 대한', '환율에 대한'과 같이 논문이 다루고 있는 주제에 대한 정부나 다른 국가, 대중들의 인식이나 행위를 언급하기 위해 주로 사용되었으며 '통하다'의 경우에는 '비교를 통해', '발언을 통해', '요인분석을 통해'처럼 논문의 주장의 근거를 제시하기 위해 사용되며 사회과학 분야에서 객관적 자료가 중요한 근거가 됨을 보여준다. '찾다'가 고빈도로 사용된 것은 공학, 자연과학 계열과 마찬가지로 그 맥락은 크게 달랐다. '그 근거는 양비론적 입장에서 찾을 수 있다', '이유를 국가의 존재에서 찾고 있다', '방안을 찾다'와 같은 맥락에서 사용된 '찾다'는 주장의 근거를 나타내는 것이며 주로 수학적 맥락에서 '찾다'가 사용된 공학, 자연과학 계열과는 다르다.

마지막으로 어문학, 인문과학 계열에서 주로 사용된 동사는 '보이다', '들다', '읽다', '찾다', '쓰다', '알다', '다루다', '살다', '먹다' 등이다. '보이다'는 주로 어학분야에서 '차이를 보이다', '예외를 보이다', '오류를 보이다' 등으로 사용되어 연구대상에 대한 객관적인 시각을 드러낸다. '들다'는 '예를 들어', '예를 들면' 등으로 사용되는데 참고문헌의 중요성이 큰 어문학, 인문과학 분야의 특징을 잘 보여주는 동사이다. '읽다'와 '쓰다'는 모두 책이나 문학과 관련된 동사인데 문학 분야의 논문은 물론이고 인문과학 분야의 문헌정보학에서도 책과 관련된 논문이 많기 때문에 '읽다'와 '쓰다'가 모두 고빈도로 나타났다. '찾다'의 경우 다른 계열에서도 많이 사용되었지만 어문학과 인문과학 계열에서는 추상적인 의미와 주로 결합된다. '자아를 찾다', '의미를 찾다', '본질을 찾다' 등이 그 예시이다. 다른 계열들에 비해 연구분야가 추상적인 어문학과 인문과학의 특징을 잘 보여준다. '다루다'는 공학, 자연과학 분야와 비슷하게 어떤 구체적인 연구대상을 언급하기 위해 사용된다. '소설을 다루게 된다', '유교적 황금률과 향약에 관해 다루었다', '이 책이 다루고 있는'에서처럼 주로 연구의 배경이 되는 텍스트나 주제, 이론을 포함하여 사용된다. '살다'와 '먹다'는 다른 계열에서는 빈도가 높지 않은 동사들인데, 인간의 삶을 주로 연구하는 어문학과 인문과학 분야에서는 다른 계열들에 비해서 '살다' '먹다'와 같은 우리의 삶과 밀접한 동사들이 주로 나타난다. "일본인"으로 살다간, '구경꾼을 모아 살고 있었다' '음식을 먹는 행위를 통해', '숙을 먹고 사람이 된 곰처럼'과 같은 문장들은 어문학과 인문과학 계열만의 특징이다.

4. 제목 분석

이번 연구에서는 지금까지 연구의 대상으로 사용한 계열별 300개 논문의 제목으로 별도의 코퍼스를 구축하여 계열별로 논문 제목 구성에 어떠한 차이가 있는지 살펴보았다. 관측 결과 계열과 관계없이 서술형 종결어미 사용을 지양하는 경향이 나타났고, 이에 따라 관형격 조사 '의', 접속조사 '와/과', 명사형 종결 형태가 빈번히 사용됨을 알 수 있었다. 이처럼 논문 제목은 핵심어에 해당하는 명사들이 조사의 도움을 받아 유기적으로 연결되어 있는 구조를 따른다. 그러나 이들 단어들의 구체적 사용 양상은 문/이과에 따라, 혹은 계열에 따라 차이가 있었다.

	인문학	어문학	사회과학	의약학	자연과학	공학
연구	68	66	59	16	55	62
-적(的)	72	59	67	75	48	26
와/과	161	141	134	96	83	50

및	9	11	36	153	56	69
의	291	332	228	307	257	173
에	80	109	110	139	135	88
을/를	86	34	55	50	90	135
공백 포함 길이(자)	35.8	24.2	29.3	40.0	34.6	30.8
공백 제외 길이(자)	28.9	19.5	22.6	30.9	27.1	23.6

표 2 : 제목에서 빈번히 사용된 단어 목록 (단위 : 회)

문/이과 분석

대체로 이과 논문의 길이가 길게 나타났다. 가장 길었던 의약학(40.0자) 이후에 인문과학(35.8자)이 등장하기는 하지만, 이를 논외로 하더라도 이과 논문에서 긴 학술 용어를 빈번하게 사용하여 제목 길이가 길었던 것으로 생각된다. 제목의 길이는 사용된 핵심어의 길이에 많은 영향을 받는다. 그러나 본 연구에서는 사용된 핵심어의 분포에 대한 분석은 논외로 하기 때문에 주로 문법적 기능을 하는 단어에 초점을 맞추었다.

먼저 눈에 띄는 사용 양상은 '및'과 '와/과'이다. 둘 다에서 최고 순위를 기록한 의약학을 제외하면, 전체적으로 문과에서는 후자를, 이과에서는 전자를 선호하는 경향을 보였다. 문법적 차원에서 접근하면 '및'은 부사이고 '와/과'는 조사이기 때문에 띄어쓰기에서 차이가 발생하나, 기본적인 기능은 동일하다. 그럼에도 이러한 사용 양상의 차이가 존재하는 이유는 두 어휘가 주는 인상이 다르기 때문일 것으로 추측된다. 이에 대해서는 추가적인 분석이 필요하다.

한편 부사격 조사 '에'는 활용 양상이 다양하게 나타났다. 인문과학과 어문학에서는 '~에 대한' 혹은 '~에 대하여'와 같은 표현이 빈번하게 드러났고, 사회과학과 의약학에서는 '~에 미치는 (영향)'이, 자연과학과 공학에서는 '~에 따른'이 특징적으로 등장한다.

목적격 조사 '을/를'은 계열과 관계없이 '~을/를 이용한', '~을/를 위한'이 많이 사용되었고, 인문과학에서는 '~을/를 중심으로'가, 사회과학에선 '~을/를 통한'이 특징적으로 드러났다.

인문과학

인문학 분야 논문의 제목은 공백이 있을 때 평균적으로 35.8, 공백 없이는 28.9자의 길이로 나타났다. 인문학 분야에서는 300개의 논문 제목 중 291개로 소유격 '의'가 많이 사용되었고 이는 연구 분야에 대한 한정, 설명을 할 때 사용된다. 실제 제목에서는 다음과 같이 나타난다.

'한국에서의 서양사 연구 한국의 서양사학과 근대성의 인식'
'과학기술인류학과 자연의 정치; 문화상대주의와 총체성을 넘어서'

그 다음으로 '와'와 '과'라는 접속 부사가 많이 쓰이는데 각각 65개와 96개로, 연구 주제를 나열하거나 그 관계에 대해 설명하기 위해 많이 쓰이는 것으로 보인다.

'과학기술인류학과 자연의 정치; 문화상대주의와 총체성을 넘어서'
'중국의 소비자 권리와 국가 소비자 운동의 전개 과정과 의미'

위와 같이 제목 한 개에도 여러 번 나타나는 경우도 많이 있다. 그리고 조사 '에'는 80 번 나타

났는데 뒤에 '대한', '대하여'와 함께 나타나는 경향이 있었다. 주제에 관한 어떤 연구를 했는지 나타낸다. 실제로는 다음과 같이 나타난다.

‘근대이행기’의 민중의식 ‘근대’와 ‘반근대’의 너머 —토지소유 및 매매관습에 대한 인식을 중심으로’

‘정치인류학에 대한 일 고찰 국가연구의 중요성 확대’

‘인터뷰에 대한 언어인류학적 성찰 정체성 관련 인터뷰 자료를 중심으로’

‘연구’라는 단어는 총 68번 출현했는데 직접적으로 논문 제목을 나타내거나 연구 분야를 나타낼 때 사용한다.

‘한국사회의 기본소득 논쟁에 대한 기독교 윤리적 실천방안 연구 - 기독교사회복지실천을 중심으로’

‘불교학 연구에 있어서 소위 문헌학적 연구의 문제점과 전망’

‘민속학 연구와 대중적 한국문화론’

‘으로’는 총 78번으로, 연구 주제에 대한 부연 설명에 많이 쓰이는데, ‘중심’이라는 어휘 앞에 함께 나타난다. ‘중심’은 65번 나타났다. 또한 ‘을’, ‘를’ 조사는 ‘중심’ 앞에 오며 각각 51, 35회 나타났다. 실제로는 다음과 같이 나타난다.

‘한국 민속학의 방향에 대한 단상 -근대화와 민간신앙의 쇠퇴를 중심으로-’

‘마을숲의 의미와 기능에 대한 고찰 -원주 성황림을 중심으로-’

또한 적(的)이라는 어휘는 72번 나타난다.

‘독한사전에서의 속담의 사전학적 처리에 대해서’

‘중한사전(中韓辭典)의 용례에 대한 사전학적 검토’

위와 같이 연구분야에 대해 한정하거나 설명할 때 많이 사용 되는 경향이 있었다.

어문학

명사 ‘연구’와 조사 ‘의’의 사용이 두드러진다. ‘의’를 비롯해 와, 과 등의 조사가 고빈도로 나타나는데 이는 논문 제목에 특수문자(괄호, 따옴표 등)가 조사 앞에 위치하여 조사가 분리된 것이다. 이를 통해 어문학 자료에 특수문자가 많이 나타남을 확인할 수 있다. 또한 원자료를 통해 확인해 본 결과 논문 제목 마지막에 사용되는 명사 중 고빈도로 나타난 것은 연구(66회), 고찰(22회), 분석(15회), 의미(9회), 양상(8회), 비교(8회), 효과(7회), 특성(5회), 변화(4회), 문제(3회), 영향(3회) 등이 있었다. 대부분의 제목들이 명사로 끝나는 명사구의 형식이었는데, 연결어미로 끝나거나 종결어미로 끝나는 경우도 일부 있었다. 예시는 다음과 같다.

‘문어몽골어의 소위 전통전사에 대하여’

‘의외성(mirative)에 대하여’

‘한국어 대용 표현의 분포와 범주적 정체성 접속 기능을 중심으로’

‘평양에선 우리 광장을 어떻게 볼까’

이 중 종결어미로 끝나는 것은 '평양에선 우리 광장을 어떻게 볼까'뿐이어서 논문에서 종결어미로 끝나는 문장을 제목으로 제시하는 것은 매우 이례적인 것임을 알 수 있다. 연결어미로 제목을 끝낼 경우 '대하여(대해서)'가 가장 일반적으로 사용되며, 그 외에 '까지', '중심으로'도 사용된다는 것을 알 수 있다. 또한 부제목을 사용한 경우가 있었는데 예시는 다음과 같다.

'아동문학을 활용한 초등 영어교육의 효과: 메타분석'

'한국 중등 학생들의 영어학습 동기모형과 변화: 종단연구'

'초등학교 영어교사의 전문성 평가도구 개발: 정의적 요인을 중심으로'

부제목을 ':'으로 구분하여 제시하였으며 부제목은 하나의 명사로 되어 있거나 연결어미로 끝난다. 제목에 비하여 연결어미를 사용하는 빈도가 상대적으로 높은 것을 알 수 있다. 또한 이 논문들은 하나의 학술지에 실린 것이기 때문에 부제목이 해당 학술지의 특징일 수 있다.

또한 하나의 명사구로 만들기 위해서 관형어가 많이 사용되었는데 주로 사용된 표현은 '대한(29회)', '나타난(27회)', '관한(20회)', '본(11회)' 등이었다. 대부분 연구 주제를 나타내기 위해 쓰였으며 이 중 '나타난'은 연구대상이 된 계층이나 텍스트를 제시하기 위해 사용되기도 하였고(일본인 여자 대학생 대화에 나타난「思う」의 사용 양상) '본' 역시 비슷한 용법으로 쓰였지만 배경이 되는 이론을 제시하기 위해 사용되기도 하였다(언어접촉의 관점에서 본 현대일본어의 통사 변화).

어문학 분야에선 '및'이 비교적 적게 사용된다. '및'은 '와/과'와 거의 동일한 의미를 갖는데 어문학 분야에서 '및'은 11회, 와/과는 141회 사용되었다.

사회과학

사회과학 제목 코퍼스에서는 '-적(的)'을 이용한 표현과 '연구'에 대한 직접적 언급이 두드러지게 나타났다. 총 300개 제목 코퍼스 중 '-적(的)'은 67회, '연구'는 59회 집계되었다. 이 외에도 '영향'이 37회, '분석'이 23회, '효과'가 11회 집계되었다.

'-적(的)'은 '그 성격을 띠는', '그에 관계된', '그 상태로 된'의 뜻을 더하는 접미사이며, 범주를 나타내는 단어(예:정치, 사회, 경제)와 결합하여 주제를 한정지어주거나, 방법(예:이론, 다면, 심층)을 나타내는 단어와 결합하여 저자의 관점을 드러내는 역할을 한다. 사회과학 계열답게 '사회적'이라는 표현이 총 11회로 가장 많이 집계되었는데, 이외 예시에서는 5회 이상 등장한 표현이 없을 정도로 다양하게 분포하고 있었다. 이러한 다양성은 그만큼 사회과학 계열이 다루고 있는 주제 범위가 광범위하다는 것을 보여준다.

- 1회 등장 : 감정적, 경험적, 관계적, 군사적, 기능적, 기본적, 다면적, 방법론적, 비판적, 상대적, 시대적, 실증적, 심적, 역사적, 외교적, 이론적, 이상적, 자기결정적, 재난적, 전략적, 문화적, 체계적, 측정학적, 통시적, 학문적, 헌법정책적, 종단적, 내재적, 실천적, 변혁적, 마케팅적, 외재적
- 2회 등장 : 개념적, 공적, 심리적, 윤리적, 정치적, 한국적, 현대적
- 3회 등장 : 탐색적
- 4회 등장 : 경제적

‘연구’는 사회과학 코퍼스에서 59회 사용될 만큼 주요한 키워드이다. 이는 한편으로는 1)‘연구’로 끝맺는 논문 제목이 많기 때문이기도 하지만, 한편으로는 2)연구 자체를 연구 대상으로 삼는 메타 연구가 활발하게 이루어진다는 뜻이기도 하다.

‘미국의 대북 군사 강압 성공에 대한 연구’

‘데이터기반의 문제해결을 통한 중소기업 원가절감 사례에 관한 연구’

‘초등학생 대상 EBS 영어 방송 콘텐츠 활용의 효과- 메타분석을 통한 연구 통합’

‘국내 영어교육분야 실행연구 동향 분석’

다른 코퍼스에서 주로 1)에 해당하는 ‘연구’의 쓰임이 많다는 점을 감안해볼 때 2)의 메타연구는 사회과학 코퍼스에서 두드러지는 방법론이라고 할 수 있다.

의약학

의약학 제목 코퍼스에서는 ‘및’을 이용한 키워드의 나열과 동사 ‘미치다’의 활용형인 ‘미치는’을 이용한 표현이 특징적으로 나타났다. 300개 논문 제목에서 ‘및’과 ‘미치는’은 각각 153회, 55회 관측되었다.

‘및’은 문장에서 같은 종류의 성분을 연결할 때 쓰는 말로, 개념어의 나열이 빈번한 논문 코퍼스에서 많이 쓰이는 어휘이다. 특히 나열하는 단어가 셋 이상일 경우에 반점을 사용하지만 두 개인 경우에는 ‘및’을 사용하는 경향이 뚜렷하게 나타나고 있다. 이는 ‘-와/과’를 우선적으로 사용하는 문과 코퍼스와 대조적인 부분이다. 의약학 코퍼스에서 ‘-와/과’는 각각 29, 67회 사용되었다.

‘전공의 교육의 현주소 및 나아갈 방향’, ‘외과 보험 현안 및 개선방향’

‘영향이나 작용 따위가 대상에 가하여지다. 또는 그것을 가하다.’라는 사전적 정의를 가지고 있는 ‘미치다’는 특정 시술이나 프로그램이 실제 환자나 관찰 대상에게 어떠한 효과를 불러일으키는지를 주된 관심사로 삼는 의약학 논문의 실험적 방법론을 가장 직관적으로 잘 드러낼 뿐 아니라 실제 논문 제목에서도 가장 빈번하게 등장하는 용언이다. 물론 절대적 빈도수에 있어서 의약학 코퍼스에서 사용된 ‘미치는’이 가장 많은 것은 아니지만, ‘미치는’이 사용되는 환경이 한정적이며 대체로 예측가능하다는 점은 의약학 코퍼스의 독특한 점이기도 하다.

실제로 ‘미치는’이 사용되는 대부분의 경우 다음과 같은 문형을 발견할 수 있었다.

문형 : A(및 B)이/가 C에 미치는 영향(효과)

예시 : ‘한국 정맥주사간호사의 업무수행도, 조직지원인식 및 개인대처자원이 직무만족에
미치는 영향’

‘미치는’은 총 55회 중 부사격 조사 ‘-에’와 54회 결합하였고, ‘영향’ 및 ‘효과’와 53회 결합하였으며, 대부분이 주격조사 ‘-이/가’와 함께 등장하였다. 이처럼 ‘(주어)가 (대상)에 미치는 영향(효과)’문형이 거의 예외 없이 유사한 형태로 등장한다는 점은 다른 코퍼스의 양상과 대비되는 의약학 제목 코퍼스의 특징이라고 볼 수 있다.

자연과학

다른 분야와 마찬가지로 ‘연구’가 가장 고빈도로 나타났다. 제목 마지막에 사용된 명사로는 연구, 분석, 특성, 비교, 추정, 탐색, 예측 등으로 나타났다. ‘연구’는 연구 주제를 표현하기 위해 사용되었으며(반응기에서의 반응 최적화 연구), ‘특성’은 연구 결과로 얻어진 특성을 제시하기 위해 사용되었다(학교 실내환경에서 공기중 미생물의 분포 및 특성). 또한 연결어미나 종결어미로 끝나는 제목들이 있었다.

‘극한과 나뭇샘 연산은 교환이 가능한가’
‘표본 개념에 대한 고찰 역사적 분석을 중심으로’

또한 특이하게 목적격조사 ‘을’을 사용한 뒤 동사 어간만으로 끝나는 제목도 있었다.(요양보호사의 특성에 따른 직무만족에 영향을 미치는 요인을 분석)

자연과학 분야에서 사용된 관형어들은 다음과 같았다. ‘대한, 이용한, 관한, 따른, 위한, 활용한, 미치는, 나타난’. 이 중 ‘이용한’과 ‘따른’, ‘위한’, ‘미치는’이 특징적이다. ‘이용한’은 주로 함수나 이론 등을 적용하여 설명하는 논문(극단값 이론을 이용한 운영리스크 심각도 분포의 추정)이나 재료, 연구 대상을 언급하기 위해(유전자변형 다분화능 정원줄기세포를 이용한 키메라 생쥐의 생산) 사용되었으며 ‘따른’의 경우 교육과정의 개정년도를 언급하여 연구 대상을 명확히 하거나(2015 개정 교육과정에 따른 수학기 교과서의 정적분 정의에 대한 대안) 어떤 두 연구대상 간에 인과관계가 있을 때(기초자산의 추세 여부에 따른 헤지변동성의 결정에 관한 연구) 등에 사용되었다. ‘위한’의 경우 연구의 구체적인 목적을 언급할 때 사용되었다.(서울시 물순환 장기 종합계획을 위한 강수량 예측 분석) 마지막으로 ‘미치는’은 연구가 독립변인과 종속변인을 포함한 과학적 연구라는 것을 보여준다(다양한 약물과 영양제가 장내세균에 미치는 영향). 이러한 관형어들이 자연과학 분야의 연구 특성을 잘 보여주는 것으로 생각된다.

자연과학 분야에서 ‘및’은 56회 사용되어 다른 분야보다 상대적으로 ‘및’이 많이 사용되었다.

공학

공학 코퍼스에서도 ‘및’이 빈번하게(69회) 사용되었다.

‘3D 프린터의 활용 및 현황 고찰’
‘선박의 저항 및 자항성능 해석을 위한 수치기법 개발’
‘대장균의 DNA 추출 방법 및 LacZ gene의 정량 방법 비교’

‘-을(를) 이용한’ 구문이 자주(67회) 사용되었다. 이를 통해 해당 연구에서 어떤 방법, 장치, 특성 등을 이용했는지 명시했다. 비슷한 맥락에서 ‘이용한’ 대신 ‘-을(를) 통한(10회)’, ‘-을(를) 사용한’(6회), ‘-을(를) 활용한’(6회)을 쓰기도 하였다.

‘Kinect와 Unity3D를 이용한 체감형 3D 가상현실 재활치료 시스템’
‘Poly 고분자 전해질을 이용한 불균질 바이폴라막 제조 및 물분해 특성’
‘데이터마이닝 알고리즘을 이용한 제품수명주기 예측’
‘ATmega128 마이크로 컨트롤러를 사용한 스마트 선풍기 속도제어’

‘연구’라는 단어로 종결한 경우는 총 62회이며, 그 중 ‘-에 관한 연구’는 33회를 차지했다. 이때는 연구 대상이 제목에 명시된다.

‘아치구조물의 형상비와 하부구조에 따른 지진응답특성에 관한 연구’
‘GPSINS센서 융합을 이용한 고정밀 위치 추정에 관한 연구’
‘라인 트레이서 로봇의 주행 알고리즘에 관한 연구’
‘유해화학물질 누출 사고에 대한 안전관리시스템 개발에 관한 연구’

‘분석’이 총 43회 언급되었다. 대상의 특성, 패턴, 실태, 요인 등을 분석하는 연구도 활발히 진행됨을 알 수 있다.

‘LANDSAT 영상을 이용한 세종특별자치시의 도시화와 열섬현상 분석’
‘Modified Hummers’ Method를 통한 Graphene oxide의 제조와 성질 분석’
‘R 프로그램을 이용한 한국프로야구의 승리요인과 패턴 분석’
‘유연 복합재료 프로펠러 제작개선 및 성능 분석’

‘-을(를) 위한’ 구문이 총 30회 사용되었으며, 이를 통해 연구 목적이 분명하게 제시된다.

‘무선 전력 전송 효율 향상을 위한 자기 공진코일 시스템’
‘스마트폰 어플리케이션 개발을 위한 모바일 플랫폼 비교 및 평가’
‘자외선 차단제 개발을 위한 이산화티탄 함유 폴리머 나노재료의 제조’
‘중금속으로 오염된 토양정화를 위한 토양세척기술 평가’

그 외에도 ‘시스템’(19회), ‘개발’(16회), ‘기술’(14회), ‘특성’(13회), ‘평가’(12회), ‘동향’(11회), ‘설계’(10회), ‘영향’(9회), ‘현황’(9회), ‘예측’(8회), ‘방법’(7회), ‘비교’(7회) 등 공학의 다양한 연구 방법론에 대한 어휘들이 언급되었다.

‘-에 따른’, ‘-이(가) -에 미치는 영향’은 각각 12회, 10회 등장하였다. 이는 특정 조건이나 방식 등에 의한 결과 혹은 결과 간 차이에 대한 연구 방식이다.

‘평직 CFRP 적층복합재료의 섬유배열각도에 따른 파괴강도 예측’
‘선풍기의 날개 수에 따른 공기 유동해석’
‘유리섬유 보강재의 적층구성, 제작공법과 수지종류에 따른 구조강도 특성에 관한 연구’
‘Foul Release 형 방오 도료가 선박의 속도 성능에 미치는 영향 연구’
‘TiO₂ 나노 입자의 크기와 결정 구조가 염료감응형 태양전지의 광전 효율에 미치는 영향’

5. 서론/결론 분석

이번 분석에서는 수집한 코퍼스를 바탕으로 논문에서 자주 나타나는 텍스트 표지를 찾고자 한다. 선행연구 중 유해준(2017)은 인문사회 분야의 논문들을 대상으로 말뭉치를 구축하여 학술 논문 텍스트의 전형적인 특성을 텍스트 표지를 통해 밝히고자 했다. 여기서 ‘표지’는 글 내용의 전개를 담당하는 담화 표지에 초점이 맞춰져 있다. 이 연구에 따르면, 학술 논문은 세부 전공 영역 별로 구조 및 문체상의 특징이 다르므로 각 영역 별 분석이 필요하지만, 이러한 연구의 시작으로 국

어학 및 한국학 소논문 10편의 서론으로 분석 대상을 한정했다. 본 연구에서는 선행연구의 연구 대상 범위를 더 넓혀 수집한 6개의 계열 논문의 서론과 결론을 분석함으로써 논문 텍스트의 특징을 알아보고자 하였다. 본론은 서론과 결론에 비해 길이가 길고, 논문의 주제에 따라 내용 및 형식이 다양하기 때문에 주어진 시간 안에 자세한 분석이 어렵다고 판단하여 제외했다. 또한 유해준(2017)은 10편의 서론으로 말뭉치를 구축하여 한마루 프로그램을 이용해 고빈도 어절들을 추출해 분석하였으나, 10편이라는 적은 수의 논문으로는 antconc를 이용했을 때 유의미한 고빈도 어절들이 추출되지 않았다. 10편 이상으로 분석 대상을 확장하기에는 수집한 코퍼스가 서론, 본론, 결론이 따로 나누어져 있지 않아 시간 상 한계가 있었기 때문에 본 연구에서는 직접 논문을 읽으며 전형적인 요소를 찾아내고 기술하는 방식을 이용했다. 본 연구를 통해 실제 논문 텍스트에서 일반적으로 나타나는 표지가 무엇인지 살펴보고자 한다.

어문학

어문학 분야에는 서론, 결론에 대해 상대적으로 다양한 표현이 사용된다. '서론/결론'이 가장 많이 사용되는 용어이기도 하지만 논문에 따라 '서언/결언', '들어가며/나가며', '들어가는 말/나오는 말'이라는 용어가 사용되기도 한다. 어문학 분야 논문 서론에 공통적으로 드러나는 가장 큰 특징은 연구대상과 연구방법을 제시하고 선행연구를 검토하는 것이다. 선행연구를 다룰 때에는 대부분 그 한계를 함께 지적하여 논문의 필요성을 강조한다. 이는 다른 분야와 크게 차이가 없다. 각 요소별로 사용되는 어휘 및 문장의 유형은 다음과 같다.

연구대상)

이 연구에서는 ~에 대한
본 논문은 ~살펴보고자 한다.
이 글은 ~ 다루어보고자 한다.
본 연구에 포함하여 살펴보고자 한다.
탐구해보고자 한다.
선행연구와 한계점
지금까지~이루어져 왔다.
~시도되어 왔다.
~한계가 있기도 하다.
~사실이 오래전부터 지적되어 왔다.
연구가 필요하다.

연구방법)

이를 위해 ~ 활용하여 ~ 분석을 수행하였다.
본 연구는 ~ 분석하여 ~을 수행
본 논문은 ~가 어떻게 ~되었는지 ~을 통해 살펴보고자 한다.
작업을 진행하고자 한다.
살펴보고 ~어떻게 반영되었는지를 분석함으로써 ~ 제시하고자 한다.

어문학 분야 서론에서 특징적인 부분은 다음과 같이 연구 주제에 대한 배경설명이 있는 경우가 있다는 것이다.

이동의 문제는 오늘날 사람들의 삶과 사회에 결정적인 요소이다. 사람의 물리적인 이동은 전 지구 상에 걸친 '생활 방식'이 되었다.

또한 문학 분야 논문의 경우 논문의 시작이 문학 작품의 인용인 경우도 있었다. 드물게 서론에서 이어지는 차례를 소개하는 경우도 있었으며 연구목적은 제시하는 논문이 종종 있었으나 다른 요소들에 비해 양이 적어서 서론의 필수적인 요소는 아닌 것으로 보인다.

어문학 분야 결론에 주로 나타나는 요소는 논문 내용 요약과 의의 제시였다. 이는 결론의 거의 필수적인 구성으로 보인다. 각 요소별로 자주 사용된 어휘나 문장 유형은 다음과 같다.

요약)

본 연구에서는 ~ 살펴보았다.

본 논문에서는 ~ 측정하였다.

그 결과를 요약하면 다음과 같다.

본고에서는 ~ 고찰하였다.

본 논문에서는 ~ 살펴보았다.

의의 제시

~점에서 연구의 의의를 갖는다.

이 연구를 통하여 ~ 알 수 있다.

이를 통해 ~ 기능할 수 있을 것이라 기대한다.

~ 제시한 것이라 할 수 있다.

이 글의 시도는 ~와 그 의미를 같이 한다.

결론에서는 그 외에도 향후 연구 방향을 제시하거나 논문의 한계를 제시하는 경우도 있었으나 필수적인 요소로 보이지는 않는다. 어문학 분야만의 결론의 특징은 찾아볼 수 없었다.

인문학

인문학 분야에서는 기존에 조명받지 못했던 부분에 대한 소개를 하거나, 기존 연구에 대한 비판, 새로운 연구의 필요성에 대해 분석하는 글이 많다. 또한 인간의 삶과 사회에 도움을 주는 방향으로 연구를 했다는 점에서 의의를 찾는다. 그리고 자신의 연구에 대한 한계, 개선점을 밝히기도 한다. 서론과 결론은 각각 '1. 서론', '4. 결론'과 같이 나타나기도 하지만 어문학과 비슷하게 '머리말'과 '맺음말', '들어가는 말'과 '나오는 말' 등 전공 분야에 따라 다르게 나타나기도 했다.

인문학의 서론에서는 연구의 필요성과 목적에 대해 설명을 한다. 기존의 연구에 대한 반박이나 새로운 관점의 연구 등을 위해 연구의 목적을 자세히 밝힌다. 목적을 서술하기 위해서는 '목적'이나 '조명' 또는 '하고자 한다'와 같은 표현을 이용한다. 연구의 필요성을 언급하기 위해서는 문제 의식이나 궁금한 점에 대해 밝힌다. 또한 그에 따라 예상되는 의의 또한 밝히는 경향이 있는데 '도움', '의의', '준거점' 등의 직접적인 명사를 사용한다.

목적)

그 같은 맥락에서 이의 목적은 다음과 같다.

본 논문에서는 ~ 조명해 보고자 한다.

필자는 ~ 밝혀보고자 한다.

그러므로 본 연구는 ~ 제안하고자 한다.

~ 주제에 관심을 두고자 한다.

연구의 필요성

이 글은 ~ 문제의식에서 출발했다.

왜 ~던 것일까.

본 연구에서는 ~를 지적하고자 한다.

의의)

~ 이해하는데 적지 않은 도움을 줄 것이다.

~ 의의를 가늠해 보고자 한다.

~ 하나의 준거점이 될 수 있을 것이다.

인문학 분야의 결론에서는 글을 전체적으로 요약하고, 연구의 의의와 한계를 언급한다. 어떠한 문제의식을 느끼고 그에 대해 새로운 관점으로 해석한 것에 대하여 의의를 말하는데, '기대', '의의' 등의 명사를 이용하여 서술하는 경향이 있다. 또한 자신의 연구에 대한 한계 또한 언급하는 것이 특징이다. 예시는 아래와 같다.

요약)

이 글에서 필자는 ~를 중심으로 살펴보았다.

~을 살펴보았다.

~는 중요한 기능을 담당하였다.

본 연구는 ~ 방안을 제시하였다.

~을 살펴보고자 한 것이다.

의의)

~해 줄 것을 기대한다.

~ 가졌음을 발견했다.

~ 의의는 다음과 같이 정리될 수 있다.

새로운 ~ 방안이 될 수 있을 것이다.

~기여할 것을 기대한다.

개선점, 한계)

아울러 이 글에서는 다루지 못했지만 ~

그러나 여전히 ~ 발전시켜야 하는 점들이 많다.

~ 연구에 관심을 기울여야 함을 말해주고 있다.

~ 할 과제가 있다고 생각한다.

사회과학

대부분의 사회과학 논문은 서론에서 연구 목적을 분명히 밝히고 있다. 이때 공통적으로 발견되는 것은 해당 연구 자체를 가리키는 표현으로, '본 연구에서는', '본 논문은', '본 연구는'이다. 연구 목적을 명시하기에 앞서 연구 배경 및 필요성, 문제점을 설명한 경우 '따라서 본 연구에서는', '이에 본 연구는'과 같은 표현을 사용한 경우도 다수 찾을 수 있었다. 그리고 종결 표현으로는 '~하고자 하였다(한다)', '~을 목적으로 하고 있다'와 같은 표지를 사용하는 경우가 대부분이었다. '-고자'는 어떤 행동을 할 의도를 가지고 있음을 나타내는 어미에 해당한다. 함께 동사는 '탐색하다', '분석하다', '모색하다', '알아보다', '도출하다', '파악하다', '제시하다', '고찰하다' 등이 활용되었다. 구체적인 양상은 다음과 같다.

연구 목적)

따라서 본 연구에서는 ~ 탐색하고자 하였다.

이에 본 연구에서는 ~를 고찰하고, 이를 바탕으로 ~를 분석하고자 한다.

살펴본 모든 논문의 서론에서 연구를 진행하게 된 동기를 설명하는 부분을 찾을 수 있었다. 이에 논문 주제와 관련하여 현황에 대한 분석과 함께 기존 방안의 문제점, 새로운 대안의 필요성 등을 언급하는 경우가 많았다. 이런 내용들은 복합적으로 제시되는 경우가 많았다. 먼저 '-변화하고 있다', '-한 상황이다', '-한 실정이다', '-고 있다'라는 표지가 기존 상황을 설명 하는데 사용된다. 그리고 여기에 '그러나', '-으나'와 같은 역접 어휘 혹은 '못하고 있다', '하지 않는다'와 같은 표현을 사용하여 기존 상황을 부정적으로 인식하고 있음을 드러내기도 한다. '가 문제로 주목받고 있다', '-문제들을 유발하고 있다'와 같이 직접적으로 문제점을 언급하기도 한다. 또한 기존 대안 및 연구의 한계나 부적절함을 지적하면서 '-이(가) 필요하다', '-할 필요성이 있다' 등의 표지를 통해 연구의 필요성을 강조하기도 한다.

연구 동기)

~ 시대로 변화하고 있다

~하기 위해 ~를 개발하고 제공할 필요성이 있다.

오늘날 우리가 살고 있는 ~이 이루어지고 있다.

그러나 ~ 않는 상황이다.

~ 등의 문제들을 유발하고 있다.

사회과학 연구에서는 서론에서 선행 연구를 언급하는 경우도 많았는데, 이는 기존 상황의 분석과 밀접한 관련성이 있었다. 선행 연구를 언급하는 경우는 크게 세 가지로 나눌 수 있었다. 먼저 본 연구와 관련된 선행연구의 개념이나 분석 내용을 인용하는 경우, 다음으로 연구 주제와 관련하여 선행 연구들이 무엇을, 왜, 어떻게 다루었는지를 단순히 설명하는 경우, 마지막으로 기존 선행연구의 문제점 및 한계를 언급하여 본 연구의 필요성을 역설하는 경우가 있었다. 물론 이들이 함께 제시되기도 했다.

선행 연구)

현재까지 여러 선행연구들은 ~하는 것이 대부분이었다.

하지만 ~점을 분석한 연구는 거의 이루어지지 않았다.

~ 대한 고찰이 요구되어지고 있다.

기존의 ~ 관련 선행연구들은 ~ 효과를 검토한 연구들과 ~을 평가한 연구들이 존재한다.

이와 관련된 연구는 거의 진행되지 않았다.

연구 방법은 본론에서 상세히 설명되는 경우가 많기 때문에 서론의 마지막에 간략히 제시되어 본론과 연결되는 경우가 많았다. 또한 연구 목적을 바로 앞에 설명하고 목적 달성을 위해 해당 방법이 필요하다는 것을 강조하기도 했다. 단, 목적을 언급하면서 방법도 같이 제시하는 경우도 있어 둘을 명확히 나누기 어려운 논문도 있었다. 사용된 표지로는 '본 연구에서는', '본 연구는'과 같이 연구 전체를 가리키는 말과, 의도를 나타내는 '-고자 하였다(한다)'가 대표적이었다. 함께 사용된 동사는 '분석하다', '살펴보다', '논의하다', '제시하다', '이용하다', '적용하다', '수집하다', '고찰하다', '파악하다' 등이 있었다.

연구 방법)

본 연구에서는 ~를 이용하여, ~를 분석하고자 한다.

분석방법은 ~ 분석을 하였다.

또한 ~뿐만 아니라 ~도 각각 분석을 하였다.

따라서 본 연구에서는 ~하고, ~을 고찰하고자 하였다.

이러한 연구목적 달성을 위해서 ~를 파악하며, ~를 토대로 ~하고자 한다.

결론에서 연구 내용을 정리해서 서술하는 부분은 앞쪽에서 쉽게 찾을 수 있다. 먼저 연구 목적과 방법, 과정 등을 설명하는 경우 '본 연구(에서)는', '본고에서는'과 같은 단어가 나타나는 경우가 많았다. 또한 목적을 요약하면서 '-하고자 하였다'라는 표현도 쓰였지만, 연구 과정을 설명하면서 별다른 어미 없이 '-하였다'라는 종결 표현이 더 많이 나타났다. '분석하다', '도출하다', '살펴보다', '제시하다', '파악하다', '이용하다', '수집하다' 등 서론에서도 쓰였던 동사들은 물론 '조사하다', '살피하다', '나타나다', '진행하다' 등의 다양한 동사들도 사용되었다.

연구 결과를 정리하면서 먼저 '이를 통해서 다음과 같은 결론을 도출하였다', '본 연구의 결과와 함의에 대해 논의하면 다음과 같다', '본 연구 결과를 요약하면 다음과 같다'라는 구문을 사용하고 결과를 나열하는 경우가 있었다. 이때, '첫째', '둘째'와 같은 서수 표현이나 '먼저', '다음으로', '마지막으로'와 같은 순서 표지가 나타났다. 또한 결과 제시에서 자주 쓰인 표지는 '-으로 나타났다', '~임을 확인하였다', '입증하였다', '~임을 알 수 있었다' 등이었다.

정리)

본고에서는 ~을 분석하였고 ~을 제시하고자 하였다.

이를 기반으로 ~를 소개하고자 하였다.

본 연구는 ~를 이용하여 새로운 시사점을 도출하는 데 있다.

본 연구결과를 요약하면 다음과 같다.

결론에서 해당 연구의 의의나 시사점, 함의 등을 언급하는 경우가 빈번했다. 먼저 직접적인 표지로는 '-데 의의가 있다', '본 논문의 - 시사점은 다음과 같다', '연구의 함의로는 -', '-임을 시사한다', '-라는 점에서 학술적 가치가 있다', '-를 시사하고 있다' 등의 표현이 사용되었으며, 간접적인 표현으로는 '-ㄹ 수 있을 것이다', '-것으로 사료된다.', '-을 보여주고 있다', '-을 확인하였다' 등이 있었다. 시사점 및 함의를 제시하면서 '-해야 할 것'과 같은 당위 표현을 사용하여 앞으로의 방향성을 강조하기도 하였다. 의의는 사실상 기대효과를 제시하는 것과도 관련성이 있다. 해당 연구가 어떤 긍정적인 영향을 미칠 수 있을지 언급함으로써 의의를 보여주는 것이다. 기대 효과는 '-ㄹ 수 있다', '-ㄹ 수 있을 것이다', 'ㄹ 것이라고 생각된다.', '-을 기대한다' 등의 가능성을 나타내는 표지와 함께 '기여하다', '활용하다', '찾다' 등의 동사가 사용된다.

의의 및 시사점)

따라서 ~이 조속히 시행되어야 함을 시사한다.

본 연구는 ~를 반영하여 연구주제를 선정하였으며, ~다는 점에서 학술적 가치가 있다.

~하면서 연구의 효용성을 높일 수 있었다.

본 연구결과는 ~ 하는데 기여할 수 있을 것으로 사료된다.

~ 참고자료로 활용가치가 있을 것이다.

결론에서는 연구 주제와 관련하여 향후 연구가 어떻게 진행되어야 하는지 방향성을 제시하는 부분이 나오기도 한다. 또한 후속 연구가 아니더라도, 해당 분야에서 정책적, 사회적 등의 측면에서

어떤 방법, 대안 등이 필요할지 연구자의 의견을 제시하기도 한다. 이때 '-할 필요가 있다', '-이 필요하다', '-하여야(되어야) 할 것이다'와 같은 표지와 '후속 연구'나 '향후 연구에서'라는 말이 등장한다. 또한 '반영하다', '수립하다', '이루어지다', '마련하다', '활용하다', '고려하다' 등의 동사가 사용되었다.

향후 방향)

따라서 ~를 이용한 후속 연구가 이루어져야 할 것이다.

따라서 본 연구 결과를 바탕으로 후속 연구에서는 ~를 ~할 필요가 있다.

결론의 마지막에서 해당 연구의 한계를 언급하는 논문도 다수 찾을 수 있었다. '-라는 한계가 있다', '본 연구의 한계로는-점이다', '-점은 본 연구가 지닌 한계이다', '하지만 본 논문에서는 -한계가 있다' 등의 표지가 쓰인다. 한계와 향후 방향은 이어서 제시되는 경우가 많았다.

한계)

본 연구의 한계로는 ~가 제외되었다는 점이다.

~한 점은 본 연구가 지닌 한계이다.

공학

공학 계열에서 살펴본 10편의 논문 모두에서 연구 목적을 밝히는 표지가 사용되었다. '본 연구에서는', '본 논문에서는', '본 연구는' 등을 주어로 하여 '~하고자 한다', '~하고자 하였다', '~에 목적을 두고 있다'와 같은 표지가 나타났다. 함께 쓰인 용언으로는 '분석하다', '적용하여', '정리하다', '개발하다', '측정하다', '수행하다', '검증하다', '규명하다', '제안하다', '평가하다', '파악하다' 등이 있었다.

연구 목적)

따라서 본 연구에서는 ~에 의한 ~을 분석하고자 한다.

이에 본 연구에서는 ~관련성을 파악하고자 한다.

대부분의 서론에서 연구를 진행하게 된 동기를 설명하는 부분을 찾을 수 있었다. 이때 최근의 동향이나 변화를 제시하는 경우가 많았으며, 이러한 상황에 적절히 대처하지 못하는 현황에 대한 문제점을 지적하기도 했다. 또한 연구에 대한 필요성을 역설하는 부분을 덧붙이기도 했다. 이때 '최근'이라는 단어와 함께 '-고 있다'라는 진행형을 자주 발견할 수 있었다. 그리고 '-에 많은 관심을 보이고 있다', '-를 많이 찾아볼 수 있다', '-가 활발하다', '-에 대해 큰 관심을 갖고 있다', '중요성이 증대되고 있다', '활발히 사용되고 있다', '-고 있는 추세이다', '-해 나가고 있다', '에서 널리 쓰이고 있다', '-에서 널리 사용되었다', '새롭게 각광받고 있다' 등의 구문을 통해 최근의 동향을 서술한다. 뿐만 아니라 '-에 대한 연구가 더욱 필요한 실정이다', '종래에는 ~많은 불편함을 초래하였다', '-문제가 제기되고 있다', '문제로 인식되고 있다', '-에 어려움을 겪고 있다', '-의 수립이 시급한 상황이다', '문제를 야기시키고 있다', '큰 위험으로 대두되고 있다'와 같은 표지를 사용해 기존 상황의 문제점을 언급한다.

연구 동기)

최근 ~고 있어 많은 불편함을 초래하였다.

하지만 ~에 문제가 제기가 되고 있다. 따라서 ~이 새롭게 각광받고 있다.

서론에서 선행연구의 내용을 언급하는 경우, 선행연구의 한계를 지적하는 경우가 많았고 새로운 혹은 발전된 방식의 연구가 필요함을 강조하기도 하였다. 이때 선행연구 내용을 설명할 때 사용된 표지에는 '기존의 -에 관한 연구는', '다양한 연구들이 진행되었다', '다양한 연구가 등장하였다', '기존에는 -을 연구해왔으나', '-등은 -을 확인하였다' '최근에는 -에 대한 연구가 활발히 진행 중이다', '-연구가 진행되고 있다' 등이 있었다. 그리고 한계를 지적할 때는 역접이나 부정어를 사용하여 '-이 반영되지 않았다', '-이 고려되지 않았다', '-하지 못하였다', '-못하는 경우가 있어 왔다', '그러나 -가 이루어진 연구는 부족한 실정이다' 등의 표지가 사용되었다.

선행 연구)

~을 위하여 다양한 연구들이 진행되었다.
이러한 연구들은 ~하였으나, ~하지는 못하였다.
~는 ~을 확인하였으며 ~을 확인하였다.
그러나 ~ 연구는 부족한 실정이다.

공학 계열은 실험적 증명과 과학적 방법론을 주로 이용하기 때문에 연구 방법과 관련된 표지가 다수 사용되었다. '-하도록 한다', '-을 이용하였다', '-을 사용한다', '-을 통해', '-이 사용되었다', '-을 관찰하였다', '-는 방법을 제안한다' 등을 예로 들 수 있다. 이때 이러한 방법을 사용하는 이유를 밝히는 목적으로 '-을 위하여'를 함께 쓰기도 하였다. 앞서 말했듯 '사용하다', '이용하다'라는 말이 가장 많이 나타났으며, 이외에 '설정하다', '고려하다', '계산하다', '설계하다', '제작하다' 등의 용언도 활용되었다.

연구 방법)

~를 위해서는 ~ 을 설정하도록 한다.
~는 ~으로 고려하도록 한다.
본 연구에서 수행한 ~는 ~를 이용하였다.
~을 분석하기 위하여 ~를 사용하였다.
~는 ~을 이용하였다.

결론에서는 연구 내용을 정리하는 부분을 쉽게 찾을 수 있다. '본 논문에서는', '본 연구는', 과 같이 논문 자체를 가리키는 표현이 함께 나온다. '-고자 하였다'를 써서 서론에서 언급한 연구 목적은 간략하게 언급하기도 한다. 그리고 '분석하였다', '-는 것으로 나타났다', '-는 것으로 판단된다', '-를 수행하였다', '계산하였다', '측정하였다', '-을 이용하여', '-을 평가하였다', '-한 것으로 사료된다', '설계하였다', '-을 확인하였다', '-을 파악하였다', '-을 제안하였다', '-을 보여주었다', '-을 검증하였다'와 같은 표지가 등장했다.

정리)

본 연구에서는 ~ 을 분석하고자 하였다.
~을 분석하였다.
~하는 것으로 나타났다.
이를 통해 ~ 것으로 판단된다.
본 논문에서는 ~을 높이기 위해 ~을 설계하였다.
값을 변화시키며 ~를 찾았고, ~을 확인하였다.

공학 계열의 결론 부분에서는 본 연구가 가지는 의의를 밝히는 표지가 나타난다. '본 연구에서',

‘본 연구는’, ‘본 논문에서는’과 ‘라고 할 수 있다’, ‘-는 점에서 의미가 있다’, ‘-는 장점이 있다’, ‘-을 확인할 수 있었다’, ‘-을 확인하였다’, ‘-는 것을 가능케 했다’와 같은 표지가 등장했다. 사회과학과 비교했을 때, ‘시사점’이나 ‘함의’을 직접 언급하는 경우는 드물게 나타나는 것으로 보인다.

의의)

본 연구는 ~ 이용하여 분석을 진행하였다는 점에서 의미가 있다.
~적용할 수 있다는 장점이 있다.

결론에서는 연구 결과가 어떤 기대효과를 불러올 수 있을지 언급하는 부분이 나타나기도 한다. ‘-할 것으로 예상된다’, ‘-하는데 적용될 수 있다’, ‘-에 활용될 수 있을 것이다’, ‘할 것으로 기대한다’, ‘-의 가능성을 열었다’, ‘-으리라 예상된다’, ‘-에 적용하고자 한다’ 등의 표현이 사용되었다. 주로 적용 및 활용 방안에 대한 가능성을 기대효과로 제시하고 있다.

기대효과)

향후 ~ 적용이 가능할 것으로 예상된다.
본 시스템은 ~하는데 적용될 수 있다.
~에도 활용될 수 있을 것이다.
~로써 활용될 수 있을 것으로 기대한다.

결론에는 향후에 어떤 연구가 필요할지 혹은 필자가 어떤 연구를 추가적으로 수행할지에 대한 설명이 제시되는 경우가 많다. 이때 ‘향후 연구에서는’, ‘후속되는 연구에서는’, ‘추후 연구로는’, ‘이번 연구를 바탕으로’, ‘차후’, ‘향후 연구로’와 같은 표지와 함께 ‘-에 대한 연구가 필요할 것으로 판단된다’, ‘-를 수행할 계획이다’, ‘-하여 -할 수 있을 것으로 기대한다’, ‘-로 확장해 나갈 것이다’, ‘-로 발전시킬 예정이다’, ‘-이 이루어질 예정이다’ 등의 표지가 나타났다.

향후 방향)

따라서 후속되는 연구에서는 ~ 를 수행할 계획이다.
향후 연구에서는 ~ 대한 연구가 더욱 필요할 것으로 판단된다.

결론에서는 해당 연구가 지닌 한계점을 제시하면서, ‘본 연구’, ‘-야 할 것이다’, ‘-할 필요가 있다’ 등의 표지가 쓰인다.

한계)

본 연구에서는 ~ 를 수행하였으나 ~ 할 필요가 있다.
하지만 ~ 되어야 할 것이다.
본 연구를 통해 ~ 을 파악할 필요가 있다.

의약학

10개 논문 중 4개의 논문에서 ‘[선행 연구 언급] - [연구의 필요성 제시] - [연구 방법 및 목적 제시]’로 이어지는 일관된 논리적 흐름을 보였다. 앞서 기술된 바와 같이, 선행 연구의 한계점은 해당 연구의 필요성과도 일맥상통하는 지점이다. 또한 이러한 한계점을 해결하는 것이 해당 연구의 목적이며, 이때의 방법론이 연구의 방법이 된다고 할 수 있다. 이 외에도 대개 서론의 구성은 기존의 상황이나 일반적 인식으로부터 시작하여 점차 연구 주제로 좁혀져 가는 논리적 구성을 따

른다.

10개 중 8개 논문에서 연구가 이뤄지기 이전의 상황을 언급하는 문장이 발견되었다. 이들 문장의 특징은 '보고', '문제', '가능성'과 같은 명사와 '-하고 있다'와 같은 현재 서술 표현이다.

기존 상황 분석)

~ 가 끊이지 않고 있다.

~또한 문제가 되고 있다.

최근에는 ~이 증가하고 있다.

~ 이 보고되고 있다.

전세계적인 고령화추세에 따라 점차 증가하고 있으며 ~

10개 중 7개 논문에서 선행연구에 대해 언급하는 문장이 발견되었다. 이는 연구자의 이름을 언급하는 경우와 '선행연구'라는 단어를 직접 사용하는 경우로 나뉜다.

선행 연구)

김영록의 연구에서~, 이유진 등은~

전진아와 이난희에 따르면 ~

선행연구에서는 ~

선행연구를 살펴보면 ~

10개 중 9개의 논문에서 연구의 필요성에 대한 문장이 발견되었다. 이들 중 4개 논문에서는 앞서 선행 연구를 언급함과 동시에 그 연구의 한계를 지적함으로써 본 연구의 타당성을 제시한다.

연구의 필요성)

그러나 ~ 연구는 거의 이루어지지 않았다.

그러나 ~ 추적 연구는 거의 없는 실정이다.

하지만 ~ 연구가 적다.

일반적으로 ~등에 대해서는 연구가 잘 되어있지 않다.

10개 중 7개 논문에서 연구 방법에 대한 문장이 발견되었다. 그중 5개 논문에서는 '본 연구'를 직접적으로 언급함으로써 독자의 주의를 환기시키고 있다. 이는 그만큼 의약학 논문에서 방법적 측면을 강조하고 있다는 점을 시사한다. 특히 의약학 논문은 글의 구조상 서론 이후에 연구 방법과 대상에 대한 상세한 기술을 통해 이를 구체화시키는 과정을 거친다.

연구 방법)

이에 본 연구는 ~ 주목하고자 한다.

본 연구는 ~ 측정하였다.

따라서 본 연구는 ~ 알아보하고자 한다.

따라서 본 연구에서는 ~ 추적 관찰하였다.

따라서 본 연구는 ~하고자 한다.

10개의 대상 논문 중 용어의 정의를 언급한 논문은 하나 밖에 없었다. 이는 의약학에서 다루는 용어들이 일반적으로 논란의 여지없이 잘 수용되고 있다는 것을 시사한다.

개념 정의)

인식이란 ~을 의미하며 ~

연구의 목적은 10개 논문 중 6개에서 발견되었으며, 대체로 글의 말미에 등장한다. 그리고 6개 중 5개 논문에서 '-하고자 하다'라는 표현이 등장하였다.

연구 목적)

따라서 본 연구는 ~하고자 한다.

본 연구에서는 ~ 확인해 보고자 실험을 실시하였다.

이에 대하여 저자들은 ~ 알아보고자 하였다.

따라서 본 실태조사는 ~하고자 한다.

이에 저자들은 ~하고자 한다.

10개 중 8개 논문에 대하여 논문의 내용을 요약하는 문장이 제시되었다. 이때 '본 연구는-'과 같은 표현을 통해 독자의 주의를 환기시키거나, '결과', '확인', '판단', '분석'과 같은 표현을 통해 연구 결과를 언급하는 경우가 가장 많았다.

정리)

본 연구는 ~ 확인하였다.

또한 본 방법을 이용하여 ~ 결과를 얻었다.

따라서 이러한 결과를 보면 ~ 판단된다.

종합하면, ~ 것으로 나타났다.

~ 본 연구는 ~ 분석하였다.

10개 중 3개 논문에서 의의와 한계를 나타내는 문장이 발견되었다. 주로 '-수 있었다'와 같은 가능 표현이 사용되어 의의가 도출되는 것과는 반대로 '-수 없었다'는 불가 표현을 통해서 한계가 드러나는 경향이 드러났다.

10개 중 2개 논문에서 '후(後)'와 '제(提)'를 이용하여 추후 연구의 방향성을 제시하고 있다.

제언)

추후에는 ~ 제언한다.

향후 ~ 연구의 필요성을 제기한다.

기대를 나타내는 표현은 10개 논문 중 6개 논문에서 발견되었으며, 이들은 대부분 '수 있다', '있을 것이다', '-으로 보인다/사료된다'와 같은 추측성 표현과 함께 결론의 말미에 제시되고 있다.

기대 효과)

~대안이 될 수 있을 것으로 보인다.

~가치가 있을 것으로 사료된다.

~자료로 활용될 수 있으며 ~

~기초자료로서 활용될 수 있을 것이다.

~어느 정도 해소할 수 있을 것이다.

의약학 논문의 경우 10개 중 7개 경우에서 결론이 한 문단 내지는 두 문단에 그쳤다. 이는 글의 구조와도 긴밀한 연관이 있다. 의약학 논문의 경우 '서론-연구 방법-연구 결과-논의/고찰-결론'

의 형태를 보인다. 이미 결론에 이르기 이전에 연구 및 실험 내용에 대한 결과 보고와 이에 따른 고찰점을 다루고 있기 때문에 결론이 전체의 요약에 가깝게 기술되고 있는 것이다. 실제로 한 개 논문에서는 아예 결론이 요약으로 쓰이기도 했다. 이처럼 결론부의 역할이 크지 않은 것은 실험적 방법론을 채택하고 있는 의약학 코퍼스의 특징이라고 볼 수 있다.

자연과학

자연과학 분야에서는 실험을 설계하고 진행한 후 그 과정과 결과를 연구 논문으로 작성한다. 자연과학 분야에서 서론은 거의 대부분의 경우 '1. 서론'과 같은 방식으로 기술된다. 결론의 경우 대부분이 '4. 결론'과 같이 나타나지만, 요약 및 제언, 결과 등의 방식으로 기술되기도 한다.

자연과학 분야의 서론에서는 진행한 연구에 대하여 그 배경과 필요성을 기술한다. 또한 실험에 대한 직접적인 내용을 설명하기 전, 실험 내용과 연구 방법을 소개한다. 연구 방법을 설명할 때는 연구에서 사용된 방법에 대한 구체적인 언급이 있고, 연구 목적을 설명할 때는 '목적'이라는 어휘를 직접적으로 사용하는 경향이 있다. 그 예시는 아래와 같다.

배경 설명)

- ~이 있었다.
- ~이 처음으로 도입된다.
- ~을 제공하였다.
- ~이 존재해 왔다.

연구의 필요성)

- ~에 대한 연구가 부족한 상태이다.
- ~는 ~에 중요한 역할을 한다.
- ~는 필수적이다.

연구방법소개)

- ~를 ~과 연계하여 살펴보고, ~를 분석함으로써 시사점을 제시하고자 한다.
- ~를 ~과 ~의 입장과 비교하고 시사점을 생각해보고자 한다.
- ~의 연구를 참조하여 ~를 분석할 ~단계 틀을 설정하였다.
- ~ 프로그램을 이용하여 ~를 예측한다.

연구 목적)

- ~에 의미를 찾는데 목적이 있다.
- ~을 목적으로 하였다.
- ~을 목적으로 한다.
- 이를 통해 ~를 검증해 본다.

자연과학 분야의 결론 부분에서는 실험에 대한 요약을 제시한다. 그 후 의의와 기대 효과, 방안을 언급하면서도, 한계점과 향후 연구의 방향을 제시해 주기도 한다. 요약을 할 때는 '본 연구에서는'이라는 말과 함께 전체적인 내용을 정리하는 경향이 나타났다. 의의를 서술할 때는 진행한 연구의 '응용', '적용', '사용' 방안을 제시했다. 자세한 예시는 다음과 같다.

요약)

이제까지 ~를 탐색하였다.
본 연구에서는 ~을 살펴보았다.
~ 결론을 얻을 수 있었다.
본 연구에서는 ~을 측정하였다.

의의)

~이 가능할 것으로 기대한다.
~에 도움이 되기를 바란다.
본 연구의 결과는 ~ 사용될 수 있을 것이다.
~ 그 응용 가능성이 매우 크다.
~ 에도 적용이 가능할 것으로 기대된다.
시사점, 향후 연구 방향 제시
~하는데 본 연구가 그 출발점이 되기를 바란다.
~에 대한 연구가 이루어지길 기대한다.
~ 몇 가지 시사점을 제시하면 다음과 같다.
~을 가능하게 할 것으로 생각된다.
차후에도 ~할 것이라 여겨진다.

한계)

~의 문제점도 고려되어야 한다.
~ 한계가 있었다.
그러나 ~은 아직 풀리지 않은 숙제이다.

이렇게 각 계열별 논문의 서론과 결론을 유해준(2017)을 참조하여 연구 동기, 목적, 선행 연구, 순서, 방법, 자료, 요약, 의의, 한계 등의 기준으로 나누어 분석해 보았다. 서론과 결론 부분에서는 계열별 글쓰기에서 말하고자 하는 바가 크게 다르지 않았다. 서론에서는 연구 동기와 목적, 선행 연구와 방법을 제시하고 결론에서는 글을 요약하고 그 의의와 한계를 밝히는 경향이 모든 계열에서 공통적으로 나타났다. 이 분석에서는 활용형이 많은데, 이들을 모두 제시하기는 어렵다는 한계가 있었다. 또한 분석 대상으로 삼은 논문의 수가 적어서 일반화시키기 힘들고, 연구 목적, 동기, 방법 등이 복합적으로 나타난 경우가 있어 분류가 명확한 분류가 어렵다는 것이 가장 큰 한계라고 할 수 있겠다.

IV. 결론

지금까지 문과와 이과의 논문 쓰기 방식의 차이점을 알아보기 위해 다양한 방법을 통해 분석을 하였다. 'AntConc' 프로그램으로 여섯 개의 분야에서 빈도수로 나타나는 단어들을 비교해 보았을 때, 상위를 차지하는 단어들은 비슷하다는 것을 알 수 있었다. 문과와 이과의 글쓰기의 논리 전개 방식에 따라 상대빈도에서 차이를 보이는 어휘들을 확인할 수 있었다. 그러나 논문의 길이에서는 문과 분야의 논문이 이과 분야의 논문보다 상대적으로 긴 것을 확인할 수 있었다. 이 단어 목록을 'Pajek' 프로그램의 네트워크로 시각화했을 때, 여섯 개의 분야에서 공기어가 높은 비중을 차지하는 것을 더 잘 볼 수 있었다. 이렇게 추출된 공기어들은 분야를 막론하고 논문 쓰기에 필요한 단어들이라는 것을 유추할 수 있었다. 'AntConc'에서 n그램을 추출했을 때, 상위 빈도 공기어들은 함께 등장

하는 경우가 많은 것을 확인할 수 있었다.

'Python'의 'Mecab 형태소 분석기'를 이용하여 문과와 이과의 논문에서 부사, 명사, 용언을 사용 빈도순으로 정리하여 비교했다. 부사의 경우 문과의 논문에서 훨씬 다양하게 사용되는 것을 볼 수 있었다. 용언은 문과의 글쓰기에서 사고와 행동에 관련된 것, 이과에서는 실험 및 관찰 조건에 관련된 것이 상대적으로 더 많이 나타났다. 명사는 논문의 주제와 연관된 키워드가 주로 나타났고, 그 외에는 계열의 특징을 나타내는 몇몇 명사에서는 차이를 보였다.

계열을 어문학과 인문학, 사회과학, 자연과학과 공학, 의학학으로 나누어 명사와 동사를 분석했다. 명사와 동사에서 상위 빈도에 공통적으로 나타나는 동사들은 학문적 글쓰기의 공통적인 특징으로 볼 수 있었고 그 외에는 계열별 학문의 특성을 나타내는 어휘들이 등장했다. 인문학과 어문학에서는 인간의 삶을 연구하고 사회과학에서는 두 변인간의 상관관계를 연구한다. 공학과 자연과학에서는 연구 대상과 수학적인 지식들이 많이 나타나며, 의학학에서는 환자와 치료에 대한 연구를 한다. 이러한 학문적인 특성에 따라, 명사에서는 상위 빈도에 대부분 주제어와 키워드에 해당하는 단어들이 나타났다. 따라서 그 계열과 학문의 특성을 확연하게 알 수 있었다. 동사에서는 상위 빈도에 논문이라는 장르의 서술 방식을 공통적으로 확인할 수 있는 동사들이 보였고, 이외에는 계열의 특징을 나타내는 동사들을 확인할 수 있었다.

문과와 이과의 제목에서는 공통적으로 '연구', '의', '에', '및' 등의 어휘가 공통적으로 많이 사용되었으며 대부분 명사 형태로 종결되는 양상을 보였다. 문과에서는 상대적으로 '적(的)'이라는 표현을 사용하여 연구의 방향이나 저자의 입장을 구체적으로 명시하는 경향과 부제목이 빈번하게 사용을 확인할 수 있었다. 이과제목에서는 연구의 방법을 직접적으로 언급하고, 실험의 요인과 결과를 보여주는 제목이 많이 나타났다. 서론과 결론을 쓸 때는 문과와 이과뿐 아니라 전공별로 표기하는 방식이 다양하게 나타났다. 서론 부분에서는 배경과 선행 연구를 설명하며 연구의 필요성을 언급하고 연구 방법을 소개했다. 결론 부분에서는 전체적인 내용을 요약, 정리하고 의의와 한계점을 서술하는데, 이과에서는 실험 결과만이 결론으로 제시되는 등 결론부의 역할이 크지 않은 것으로 보였다.

이와 같이 문과와 이과, 네 개의 하위분류 계열별 학문에서의 차이점과 공통점을 확인했다. 원래의 예상과는 다르게 각각의 학문 영역들 사이에 차이점보다는 공통점이 더 많이 관측되었으나, 연구 방식의 세밀화와 다각적 접근을 통해서 유의미한 차이점들을 발견할 수 있을 것으로 예상된다. 한편 코퍼스에서 띄어쓰기의 문제가 발생했는데, 이는 코퍼스 구축 과정에서 맡은 부분을 전 처리할 때 PDF 파일인 논문을 TXT 파일로 변환하는 프로그램의 기능적 문제 때문이다. 이 문제를 해결하기 위하여수작업으로 이루어진 부분이 많아 코퍼스의 크기를 크게 늘리지 못했던 점이 한계라고 할 수 있다. 또한 한국어정보만을 위해 영어와 숫자 정보를 일괄적으로 삭제했는데 이 때문에 조사가 상대적으로 많이 집계된 상황이 드러난다.

추측에 불과했던 문과와 이과, 계열별 글쓰기의 차이를 연구를 통해 실제로 확인할 수 있었다. 이러한 결과는 고급 한국어 학습자를 위한 학습 자료로 사용될 수 있을 것이다. 구체적으로 N-gram 분석 결과를 통해 학문적 글쓰기에서 자주 사용되는 어구의 예시와 계열별로 자주 사용되는 동사와 명사, 논문의 제목과 서론, 결론의 구조의 예시를 제공할 수 있을 것이다. 더 많은 자료들을 통해 내용을 보충한다면 이는 고급 한국어 학습자에게 안내가 될 뿐 아니라, 학문적 담화 코퍼스로 구축이 되어 향후 자동 수정, 자동 완성 등의 기술에도 활용이 가능할 것으로 기대된다.

참고문헌

박나리. 2012. '-는 것이다 구문연구'. 국어학, 65: 251-279.

최도식. 2015. '계열별 글쓰기의 특성과 교육방법적 모형 개발'. 시학과 언어학, 29: 177-205.

유해준. 2017. '한국어 학술 논문텍스트 표지 분석'. 영주어문, 37: 311-326.