4. 결론

지금까지 문과와 이과의 논문 쓰기 방식의 차이점을 알아보기 위해 다양한 방법을 통해 분석을 하였다. ‘AntConc’ 프로그램으로 여섯 개의 분야에서 고빈도로 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 등장하는 경우가 많은 것을 확인할 수 있었다.

‘Python’의 ‘Mecab 형태소 분석기’를 이용하여 문과와 이과의 논문에서 부사, 명사, 용언을 사용 빈도 순으로 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.

------계열별 분석 부분------

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. 이과 제목에서는 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 이과의 여섯 개의 하위 분류 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. 코퍼스 내에서 특히 부사, 접속사, 동사들의 등장 빈도와 다양성 등을 비교해보고 연구의 객관성을 검증해주는 통계학적 어휘(표본, 회귀분석, 모집단 등)의 사용 빈도를 살펴봄으로써 새로운 결과를 발견할 수 있기를 기대해본다.

한편 코퍼스를 구성하고 있는 어절의 개수를 측정했을 때 띄어쓰기와 영문/숫자 정보의 개입 등으로 원치 않는 결과가 유입되는 문제가 발생했다. 이는 코퍼스 구축 과정에서 맡은 부분을 전처리할 때 각기 다른 방식을 차용했기 때문이다. 따라서 완성된 코퍼스를 일괄적으로 재처리하는 과정을 통해 영문과 숫자 정보를 일괄적으로 삭제하고, 띄어쓰기를 재점검한다면 조사 결과의 신뢰도와 정확도가 향상될 것으로 예상된다.