

연구 논문의 의미 구조 기반 메타데이터 항목의 자동 식별 처리를 위한 문장 구조 분석*

Analyzing the Sentence Structure for Automatic Identification of Metadata Elements based on the Logical Semantic Structure of Research Articles

송민선 (Min-Sun Song)**

초 록

본 연구는 연구논문의 논리적 의미 구조 메타데이터 항목에 해당하는 데이터에 담겨 있는 문장의 구성에 따라 시스템에서 적절한 항목으로 자동 식별·처리될 수 있도록 하는, 문장의미론(Sentence Semantics)적 분석 방법을 제안하고자 하는 목적으로 수행되었으며, 의미 구조 메타데이터 항목 중 'Research Objectives'와 'Research Outcomes'에 해당하는 연구 논문 문장의 구조를 어절 수, 접속어 종류, 다수 출현한 단어들의 문장 내 역할, 문장에서 다수 출현한 어미 형태 등을 기준으로 분석해 정리하였다. 연구 결과, 문장들의 어절 수는 'Research Objectives'는 평균 38개, 'Research Outcomes'는 평균 212개로 나타났으며, 접속어의 경우 'Research Objectives'는 인과-순접-대등-환연/요약 관계를 나타내는 접속어 순으로, 'Research Outcomes'는 인과-대등-순접-환연/요약 관계를 나타내는 접속어 순으로 많이 출현한 것으로 파악되었다. 출현빈도가 높은 분석 대상 단어들은 각각 문장 내에서 주어, 목적어, 서술어 역할 등으로 사용되고 있었으며, '역할'이나 '요인', '관계'는 목적어나 결과 부분 모두에서 비슷한 역할을 담당하고 있었지만 '연구'는 같은 단어라도 연구의 목적 부분과 결과 부분에서 사용되는 역할에 차이를 보였다. 마지막으로 문장 내 동사의 어미는 'Research Objectives'에서 '~고자'와 '~었다', 'Research Outcomes'에서 '~었다', '~있다', '~였다'가 많이 출현하였다. 본 연구는 연구자의 학술적 이해형성을 지원하기 위해 연구논문이 담고 있는 공통된 논리적 의미를 반영한 메타데이터 요소의 자동 식별과 입력 방안을 제시하는 데 활용할 수 있는 기초 연구로서 의의가 있다.

ABSTRACT

This study proposes the analysis method in sentence semantics that can be automatically identified and processed as appropriate items in the system according to the composition of the sentences contained in the data corresponding to the logical semantic structure metadata of the research papers. In order to achieve the purpose, the structure of sentences corresponding to 'Research Objectives' and 'Research Outcomes' among the semantic structure metadata was analyzed based on the number of words, the link word types, the role of many-appeared words in sentences, and the end types of a word. As a result of this study, the number of words in the sentences was 38 in 'Research Objectives' and 212 in 'Research Outcomes'. The link word types in 'Research Objectives' were occurred in the order such as Causality, Sequence, Equivalence, In-other-word/Summary relation, and the link word types in 'Research Outcomes' were appeared in the order such as Causality, Equivalence, Sequence, In-other-word/Summary relation. Analysis target words like '역할(Role)', '요인(Factor)' and '관계(Relation)' played a similar role in both purpose and result part, but the role of '연구(Study)' was little different. Finally, the verb endings in sentences were appeared many times such as '~고자', '~었다' in 'Research Objectives', and '~었다', '~있다', '~였다' in 'Research Outcomes'. This study is significant as a fundamental research that can be utilized to automatically identify and input the metadata element reflecting the common logical semantics of research papers in order to support researchers' scholarly sensemaking.

키워드: 학술적 이해 형성, 논리적 의미 구조, 연구 논문, 메타데이터, 문장 구조 분석, 문장의미론
scholarly sensemaking, logical semantic structure, research articles, metadata,
analyzing the sentence structure, sentence semantics

* 이 논문은 2017년도 성균관대학교 박사후 교내 연수지원 사업에 의하여 연구되었음.

** (전)성균관대학교 정보관리연구소 연구원, (현)대림대학교 아동문헌정보과 조교수(songser@daelim.ac.kr)

■ 논문접수일자: 2018년 8월 19일 ■ 최초심사일자: 2018년 9월 13일 ■ 게재확정일자: 2018년 9월 17일
■ 정보관리학회지, 35(3), 101-121, 2018. [http://dx.doi.org/10.3743/KOSIM.2018.35.3.101]

1. 서론

1.1 연구 목적 및 필요성

일반적으로 연구자들의 연구 활동은 관심 주제 분야에 대한 연구 목적과 필요성에 따라 제시된 연구 문제를 해결하기 위한 가설의 설정, 연구 대상 데이터의 수집·가공, 실험 및 관찰 결과의 검증 과정의 단계로 구성된다. 연구 논문은 이와 같은 연구 활동의 결과를 연구자가 논문의 논리적 지식 구조에 따라 표현한 결과물으로써, 대체로 연구 목적과 필요성, 해당 주제와 연관된 유사 연구들에 대한 분석, 해당 연구의 실험 과정과 검증 결과의 내용으로 구성된다.

미래창조과학부(2015)의 보고서에 따르면, SCI 등재 논문만의 경우에도 세계에서 발간되는 총 논문 수가 2010년 122만여 건에서 2014년 146만여 건으로 꾸준히 증가하고 있는 추세를 보이고 있다. 또한 논문의 수가 증가하는 만큼 연구 대상 주제 분야들도 세분화·다양화되고 있으며, 서로 다른 분야 간의 공동 연구가 증가하고(The Royal Society, 2011), 연구 주제 역시 융·복합적인 특성을 많이 보이는 것으로 나타나고 있다.

이와 같이 다양한 분야에서의 연구 논문 수가 증가하면서, 연구 활동 과정에서 유사 연구를 수행한 다른 분야 연구자들의 연구 성과에 포함된 전문 지식을 참고해 활용하고, 이를 토대로 새로운 연구를 수행하는 연구자들이 분석하고 해결해야 할 정보문제 역시 점점 늘어나고 있다. 특히 유사 분야 내에서 비슷한 주제 영역을 다루는 여러 연구 성과물들의 의미적인 연관 관계를 파악하고, 그 차이점이나 유사점을 정리하는 일은 연구자들이 연구 수행 단계

에서 반드시 거쳐야 할 단계이다. 하지만 오늘날 연구 논문들은 해당 논문에 포함되어 있는 논리적 내용에 기초한 의미 구조보다는, 논문 제목을 비롯해 저자명, 초록, 목차 등 형식적인 서지사항을 토대로 한 메타데이터 시스템 내에서 관리되고 있는 것이 현실이다. 해외의 일부 저널의 경우는, 논문 투고 과정에서 저자가 초록을 작성할 때 Purpose, Design/methodology/approach, Findings, Practical implications, Originality/value 등의 내용을 직접 입력하도록 해 연구자들이 연구 수행 과정에서 의도한 내용들을 초록에서 바로 확인할 수 있도록 시스템화 하고는 있지만, 아직까지 대부분의 일반 논문 검색 시스템들은 서지 형식 기반 메타데이터를 토대로 하고 있으며, 초록 작성 역시 연구자들의 자율에 맡기고 있어 원 연구자가 의도했던 연구의 전체적인 흐름이나 논리적 내용들을 검색 결과에서 바로 확인하기 어렵다.

따라서 연구자들이 연구를 수행하는 과정에서 발생하는 여러 가지 정보문제들을 효과적으로 해결하기 위해서는 기본적으로 연구 논문 데이터를 저장·관리·제공하는 정보 시스템의 구조를 기존의 형식 기반 메타데이터 항목에 포함된 키워드 검색 지향형에서 학술적 의미관계를 파악할 수 있는 정보문제 해결 지향형으로 개선하는 작업이 필요하다고 할 수 있다. 즉, 연구 논문의 메타데이터를 연구 목적, 연구의 필요성, 연구 방법 등과 같이 논문이 담고 있는 논리적 의미 구조에 따르는 내용적인 지식 특성을 반영하는 항목으로 구성한다면, 연구 과정의 각 단계에서 발생하는 연구자들의 학술적 정보 요구를 충족시키는 데 필요한 정보를 식별하고 파악하는 데 많은 도움을 줄 수 있을 것이다.

이와 관련하여 송민선과 고영만(2015)은 연구 논문의 의미 구조와 관련된 기존 연구 분석 내용을 바탕으로 공통된 의미 구조 메타데이터 항목들을 도출하고, 융복합적 성격을 띄는 한국학 분야 학술 논문들에 포함된 저자키워드의 유형 분류 작업 및 메타데이터 항목과의 매핑 작업을 수행해 의미 구조 메타데이터 항목을 조정·정리하는 연구를 수행하였다. 또한 해당 메타데이터를 기반으로 하는 시범시스템(Korean Studies Metadata Database: 이하 KMD)을 구축함으로써, 연구자가 의도한 내용을 검색하는 데 있어 정확률을 높일 수 있고 원하는 검색 결과의 내용을 선별해 보여주기 용이하다는 연구 결과도 제시된 바 있다(송민선, 고영만, 이승준, 2016). 그러나 KMD 시스템 구축과 관련하여 각 의미 구조 메타데이터 항목에 해당하는 데이터 입력 작업 수행 시 전문 인력들이 직접 관련 논문들을 읽고 해당 항목에 대한 데이터를 입력하는 수작업 방식의 경우, 입력 시간 대비 작업 효율성과 통일성이 떨어지는 문제점이 있는 것으로 드러났다.

한국학중앙연구원의 경우도, 한국학 분야 연구자들이 한국학의 큰 줄기와 흐름을 파악해 시대 및 세부 분야별로 소외된 분야 없이 균형 잡힌 한국학 연구를 수행할 수 있도록 돕기 위해, 2013년부터 2016년까지 기존 연구 성과들에 포함된 연구대상, 시대배경, 주제, 이론적 쟁점이 무엇인지에 대한 전문가 차원의 분석 결과를 제안하는 ‘한국학 연구지형도’ 시스템 개발 사업을 추진한 바 있다. 이 사업 역시 매년 각 분야별로 데이터 입력을 위한 과제 공모를 통해 동일하게 수작업 방식으로 데이터 입력 작업을 수행하였으나 실제 입력된 데이터를 살

펴보면 각 분야 및 입력자에 따라 데이터 품질의 차이가 나타났다.

전문 인력에 의한 직접 데이터 입력 방식은, 각 메타데이터 항목에 적절한 내용을 직접 확인하고 적합성을 판단한 후 입력이 진행된다는 점에서 정확성이 높다는 장점을 갖는다. 하지만 수작업 방식의 경우, 논문을 모두 읽고 각 메타데이터 항목에 해당하는 내용을 구분해 입력해야 하기 때문에 작업 시간이 많이 소요되며, 각 논문이 담고 있는 내용과 논리, 맥락 등을 이해할 수 있는 수준의 전문인력 고용이 필요하기 때문에 투자 비용 대비 작업 시간의 효율이 떨어지고, 동일한 입력 지침과 사전 교육을 진행하고 작업을 수행하더라도 입력자의 작업 성향이나 세부 전공 분야 등에 따라서 데이터의 질적 수준 차이가 발생할 수 있다는 단점들을 배제할 수 없다.

본 연구에서는 여러 분야 연구논문에 수록되어 있는 논리적 의미 구조에 해당하는 실제 문장들을 분석하여, 해당 항목의 콘텐츠에 담겨 있는 문장의 구성 등에 따라 시스템에서 적절한 메타데이터 항목으로 자동 식별해 입력될 수 있도록 처리하기 위한 문장의미론(Sentence Semantics)적 분석 방법을 제안하는 것을 목적으로 한다. 여기에서 문장의미론적 분석 방법이란 문장 단위에서 발견되는 여러 의미들을 체계적으로 분석하는 분야로, 개별 문장 자체 의미 뿐만 아니라 문장들 사이의 의미적 관계, 문장 내 의미 구조 등을 연구하는 방식을 말한다(신준철, 옥철영, 2016).

1.2 선행 연구 분석

앞서 언급했듯이 학술 연구 논문의 논리적 구

조는 연구자의 정보 처리 및 학술적 이해 형성 과정과 맞물려 있다고 할 수 있다. 이와 관련하여 90년대 중후반 이후로 여러 관련 연구들이 진행되었는데, Kando(1997; 1999)는 연구 문헌의 문서 구조를 고려한 주제 색인에 대해 연구를 수행하였으며, Horn(2000)은 연구 문헌의 지식 구조를 기반으로 연구자의 정보 요구를 지원하는 논증지도(argumentation map)를 제안하였다. 이 외에도 수사적 의미 관계 기반의 학술지 지식구조 표현 형식 및 정보 공유 체계를 연구한 Harmsze(2000), 기초학문자료센터(Korean Research Memory, 이하 KRM)의 데이터를 분석해 연구자 중심의 연구성과물 의미검색을 위한 온톨로지 구축에 대한 연구를 수행한 유사라, 이해원, 송인석(2009), KRM 연구성과물 DB와 용어 DR의 개념 클래스 및 인스턴스를 대상으로 연구 문헌의 학술적 개념 요소를 제시한 고영만과 송인석(2011)의 연구 등은 연구 문헌이 연구자가 체계화한 학술적 의미 기반 지식 구조로 이루어져 있음을 잘 보여주고 있으며, 해당 구조에 기반한 검색 시스템을 제대로 개발해 활용할 경우 연구자들의 연구 활동을 보다 효율적으로 지원할 수 있음을 시사하고 있다.

한편, 한국어 구문의 자동 식별 분석과 처리와 관련해 문헌정보학 분야에서는 도서관에서 서비스하는 정보를 효과적으로 검색할 수 있도록 하기 위한 목적에서 문헌 내에 나타나는 단어의 출현 빈도와 출현 위치 등에 의한 자동 색인 연구가 주로 이루어져 왔으며(윤구호, 1999; 정영미, 이태영, 1982), 최근에는 인터넷에서 제공되는 정보와 관련된 데이터 마이닝이나 기계학습을 활용한 자동 색인, 분류 등의 연구가

많이 이루어지고 있다(김민철, 심규승, 한남기, 김예은, 송민, 2013; 강범일, 송민, 조화순, 2013).

웹 상에서 사람이 작성한 문서나 문장들을 실제 사람의 논리 구조나 의견과 흡사하게 분류하고 처리하고 검색해 이용할 수 있도록 하기 위한 연구들은 컴퓨터 공학 주로 많이 이루어져 왔다.

한정기, 박민규, 김준태(1998)는 한국어 특성에 맞춰 사람들이 표현하는 의견 데이터를 효과적으로 분류하기 위해 어미 변화 패턴을 활용한 문장 성분 분석 및 분류 방식을 제안하였으며, 윤성희와 백선옥(2004)은 질의응답 시스템 내에서 이용자가 질의한 자연어 문장에서 질의의도를 파악하고 해당 유형을 효율적으로 분류하고자 초점 어휘를 추출해 구문 구조적으로 관련 단어의 의미 정보에 기반해 질의 유형을 분류하는 방법을 제안하였는데, 질의 초점 어휘가 생략되었을 때에는 동의어 및 접미사 정보를 이용해 유형 분류 성능을 향상시킬 수 있는 방법도 함께 제시하였다. 김진옥, 이선숙, 용환승(2011)은 데이터 마이닝 기술을 활용하여 훈련 데이터에 빈번하게 나타난 어절을 이용해 빈도수 기반의 점수 부여 방식을 통해 웹 상의 텍스트에 나타난 저자의 의견을 긍정, 중립, 부정으로 판단하는 알고리즘을 제안하였다. 안애립, 한용진, 박세영, 남지순(2012) 역시 데이터 마이닝 기술을 활용해 사용자 의견문장을 분석하는 연구를 수행하였는데, 함께 공기하는 자질 명사에 따라 그 극성 값이 달라지는 '비결정 오피니언 어휘' 처리를 위해 '맛집' 도메인과 관련한 키워드 목록을 결정한 후 부분문법그래프 모델을 적용해 현 단계 오피니언 자동 분류 시스템의 성능을 향상시킬 수 있음을 확인하였다.

한국어 문헌 및 인터넷 상의 한국어 문서(문장)의 자동 분석, 색인, 분류 처리와 관련한 이상의 연구 결과들은 연구논문을 대상으로 한 의미 검색 시스템 개발에 있어 연관된 메타데이터 항목 내 관련 데이터들의 문장 분석에 활용함으로써 데이터 입력 과정에서 자동 식별 처리 기술을 개발하기 위한 일련의 규칙들을 도출하는 데 도움을 줄 수 있다.

하지만 학술 연구 논문의 논리 기반에 근거한 의미 구조 메타데이터 항목들을 자동 식별해 처리하기 위해 문장의미론적 접근 방법을 적용·분석한 연구는 아직까지 수행된 바가 없다. 또한 앞서 언급했듯이 본 연구는 기존의 실험적 연구들(송민선, 고영만, 2015; 송민선, 고영만, 이승준, 2016)에서 발생한 연구 단계 과정에서의 효율성 문제를 해결하기 위한 실용적인 방법론을 제시한다는 점에서 기존 유사 연구들과는 차별성을 갖는다고 할 수 있다.

1.3 연구 방법

본 연구는 의미 구조 메타데이터 항목에 입력된 데이터들의 실제 문장에 포함된 단어와 구조 등을 분석하여, 공통점과 차이점들을 제안함으로써 각 의미 구조 메타데이터 항목에 해당하는 내용을 시스템에서 자동 식별해 입력 처리할 수 있는 단서를 제공하기 위한 목적을 지닌다.

이를 위해 본 연구에서는 ① 분석을 위한 연구 논문 데이터 선정, ② 의미 구조 메타데이터 항목별 데이터 입력 작업 수행, ③ 분석 대상 메타데이터 항목에 입력 데이터의 수록 문장들에 대한 형태소 분석 작업, ④ 문장 내 단어 구

성 분석에 의한 의미 구조 파악 작업의 단계로 수행되었다.

먼저 분석을 위한 연구 논문 데이터는 한국학술지인용색인(Korea Citation Index, 이하 KCI)에서 제공 중인 학술지 주제 구분 중 '인문학, 사회과학, 자연과학, 공학, 의약학, 농수해양학, 예술체육학, 복합학'의 8가지 분야로 나뉜 대분류를 기준으로 피인용 빈도가 높은 상위 50위(2017년 6월 1일 기준) 논문 413편 중 본문이 한국어가 아닌 외국어로 작성된 14건을 제외한 399건의 논문으로 선정하였다(〈표 1〉 참조). 피인용 빈도가 높은 논문은 해당 분야 연구자들의 활용 빈도가 높다는 의미와 같기 때문에 그만큼 논문의 내용과 형식에 있어 신뢰성이 있다고 판단했기 때문이다. 피인용 빈도 순위를 50위로 제한한 까닭은 기준 순위를 높일수록 각 주제 분야별로 피인용 횟수가 동일한 순위별 논문 누적 수의 차이가 많이 났기 때문이다. 따라서 각 대분류에 속한 피인용 빈도 상위 논문들의 누적 수 차이가 비교적 적은 순위를 기준으로 결정하였다.

〈표 1〉 구축 대상 논문 데이터 수

KCI 대분류	구축 대상 논문 편수
인문학	48
사회과학	51
자연과학	51
공학	49
의약학	44
농수해양학	53
예술체육학	51
복합학	52
총계	399

둘째, 구축 대상 논문의 내용을 의미 구조 메타데이터 항목별로 구분해 입력 작업을 수행하

였다. 의미 구조 메타데이터 항목은 송민선과 고영만(2015)에서 제안하고, 송민선, 고영만, 이승준(2016)에서 시범시스템 구축을 위해 적용한 의미 구조 메타데이터 항목을 기준으로 하였다(〈표 2〉 참조). 또한 최대한 학술논문들이 포함하는 공통된 논리적 의미를 반영해 메타데이터를 구성하기는 하였지만, 세부 분야별로 논문의 논리적 의미 구조에 일정 부분 차이가 있을 수 있다는 점은 배제할 수 없다고 판단하였기 때문에 모든 의미 구조 메타데이터 항목은 필수가 아닌 선택 항목으로 두고 입력을 수행하였다. 분석을 위한 데이터는 MS Excel 프로그램에 수작업 방식으로 입력하였는데, 앞서 언급했던 입력자별 데이터의 질적 차이를 줄이기 위해 본 연구자가 단독으로 작업을 수행하였으며 원저자가 논문에 기술한 문장을 그대로 살려 데이터를 입력하였다.

셋째, 데이터 입력을 완료한 후 의미 구조 메타데이터 항목들 중 ‘Research Objectives’와 ‘Research Outcomes’에 입력된 데이터들을 중심으로(〈표 2〉 참조), 아리랑 한국어 형태소 분석기(Arirang Analyzer)를 사용해 형태소 분석을 실시하였다. 분석 대상 데이터를 위의 2종 메타데이터의 입력 내용으로 제한한 까닭은 분석 대상으로 입력한 모든 논문에서 포함하고 있는 항목이었기 때문이다. 그 외 메타데이터 항목들의 경우 제3자 입장에서 판단했을 때 연구 분야나 대상 주제, 저자의 논문 기술 방식에 따라 논문 전체에서 명백히 드러나지 않은 경우들이 있어 저자가 기술한 내용 그대로 옮겨 입력할 수 없었으므로 분석에서는 제외하였다.

넷째, MS Excel 프로그램을 사용하여 형태소 분석 과정을 거쳐 정리된 명사 형태 단어가

〈표 2〉 의미 구조 메타데이터 항목

의미 구조 메타데이터 항목	메타데이터 설명
Research Objectives*	연구 목적
Research Significance	연구의 의의
Research Limitation	연구의 제한점
Research Type	연구 유형(‘질적연구’, ‘양적연구’, ‘혼합연구’로 구분)
Research Problem / Hypotheses	연구 문제 / 가설
Research Method	연구 방법
Data	연구대상 데이터
Temporal Background	연구대상의 배경이 되는 시간적 범위
Spatial Background	연구대상의 배경이 되는 공간적 범위
Related Models	연구와 관련되거나 차용한 이론 및 모형 등
Method of Data Collection	데이터 수집 방법
Data Size	데이터 수집 규모
Analysis Method	연구 데이터 분석 방법
Research Outcomes*	연구 결과
Further Research	후속 연구 제언

※ ‘*’ 표시 항목이 최종 분석에 적용한 메타데이터 항목임.

포함된 어절 및 문장에 사용된 접속어, 동사의 어미 등을 구분·정리함으로써, 각 메타데이터 항목에 입력되어 있는 문장 내 단어의 구성 분석에 의해 문장의 의미구조를 파악하는 작업을 수행하였다.

먼저 각 메타데이터 항목별로 입력된 문장들에서 형태소 분석 과정을 통해 접속사, 조사, 어미 등을 분리한 명사 형태의 단어를 중심으로 추출해 해당 논문에서 출현한 빈도수를 기준으로 정리하였다(〈그림 1〉 참조).

다음으로 논문 1편의 해당 메타데이터 항목에서 2회 이상 출현한 단어들을 기준으로 전체 논문에서의 출현빈도를 계산해 정렬하고, 누적 출현빈도 기준 상위 50% 이상에 포함되는 단어들 중 일반적으로 어떤 주제 분야의 논문에서든 연구 내용을 기술하는 데 있어 사용될 수 있는 일반적인 의미의 단어들(예: 연구, 의의, 관계, 결과, 제공, 목적, 영향, 분석, 파악 등)이 포함되는 부분들을 각 문장의 해당 어절에 구분 표시하였다. 또한 형태소 분석 과정에서 제

외된 접속어(예: 그래서, 그리고, 그러므로, 따라서, 하지만 등)와 동사 변화에 적용되는 어미 조건 부분들(예: ~이다, ~있다, ~였다, ~했다, ~(하)고자 등)이 포함되는 내용들을 각 문장의 해당 부분에 표시하여 전체적으로 각 메타데이터 항목에서 어떤 종류의 접속어와 어미 형태, 어떤 의미의 단어들이 논문의 관련 항목을 기술하는데 주로 사용되는지를 판별할 수 있도록 하였다(〈그림 2〉 참조).

2. 의미 구조 메타데이터 항목별 문장 분석 결과

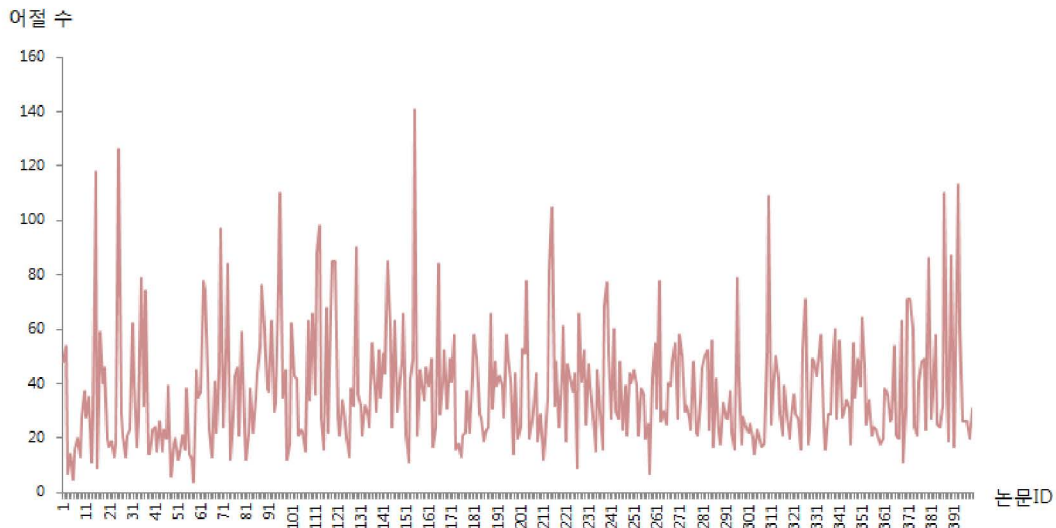
앞서 기술한 바와 같이 본 연구에서는 분석 대상 논문 전체에 입력된 ‘Research Objectives’와 ‘Research Outcomes’ 메타데이터 항목의 데이터를 대상으로 분석을 수행하였다. 주요 분석 내용은 ① 각 항목별 문장을 구성하는 어절의 수, ② 문장 간 의미 관계를 파악하는 데 중요한

연구목적	연구목적
따라서 본 연구는 포털조직의 구성원의 조직행동에 긍정적 변수라 할 수 있는 직무만족과 부정적 변수라 할 수 있는 직무스트레스 그리고 일과적 변수들 중 하나인 이직의도와의 관계를 규명하여 포털조직의 생산성을 향상시키고 동시에 효과적인 이직관리를 위한 기초 자료를 제공하는데 있다.	연구, 포털조직, 구성원, 조직행동, 긍정적, 변수, 직무만족, 부정적, 변수, 직무스트레스, 일과적, 변수, 이직의도, 관계, 규명, 생산성, 향상, 동시, 효과적, 이직관리, 기초, 자료, 제공
본 연구에서는 지역사회거주 노인들의 우울감 정도를 파악하고 노인들의 우울감에 영향을 미치는 주요 요인들을 규명하고자 한다. 이를 통하여 노인들의 정신건강 증진을 위하여 개별적 정신상태를 고려한 지역재가복지사업, 정신보건서비스 및 관련 사회복지서비스 등을 활성화시킬 수 있는 대책과 방안을 제시하고자 한다. 본 연구의 구체적인 연구목적은 다음과 같다. 첫째, 지역사회거주 노인들의 우울감 정도를 파악한다. 둘째, 지역사회거주 노인들의 종류에 영향을 미치는 사회인구학적 특성, 주관적 건강상태, 기능적 능력, 사회적 지지, 여가활동 및 생활만족도 등과의 관계를 조사한다. 셋째, 지역사회거주 노인들의 우울감에 영향을 미치는 결정요인들을 규명하고자 한다.	연구, 지역사회거주, 노인, 우울감, 정도, 파악, 노인, 우울감, 영향, 오인, 규명, 노인, 정신건강, 증진, 개별적, 정신상태, 고려, 지역재가복지사업, 정신보건서비스, 사회복지서비스, 활성화, 대책, 방안, 제시, 연구, 구체적, 연구목적, 다음, 지역사회거주, 노인, 우울감, 정도, 파악, 지역사회거주, 노인, 종류, 영향, 사회인구학적, 특성, 주관적, 건강상태, 기능적, 능력, 사회적, 지지, 여가활동, 생활만족도, 등과의, 관계, 조사, 셋째, 지역사회거주, 노인, 우울감, 영향, 결정요인, 규명
따라서 본 연구에서는 농촌 국제결혼 가정 아동의 언어발달 실태를 알아보고 언어 발달에 영향을 주는 변인들을 알아봄으로써 사회적 고독적 지원 방안을 모색하여 궁극적으로는 아동의 발달과 외국어 학습의 한국사회의 안정된 정착을 도모하고자 하며 이와 같은 목적을 수행하기 위해 다음과 같은 연구목표를 설정하였다. 첫째, 국제결혼 가정의 아동 언어능력에 대하여 수용언어와 표현 언어의 발달정도를 알아본다. 둘째, 국제결혼 가정의 언어 환경을 알아본다. 셋째, 아동의 언어발달과 어머니의 한국사회 정착을 위한 사회적, 교육적 지원 요구 사항을 알아보고 문제 해결 방안을 모색한다. 본 연구는 이러한 문제점에 착안, 기존의 다양한 문헌을 바탕으로 기업의 사회적 책임활동을 가장 바람직하게 반영할 수 있는 개념을 기초로 이를 측정할 수 있는 척도를 개발하는 것을 목적으로 하고 있다.	연구, 농촌, 국제결혼, 가정, 아동, 언어발달, 실태, 언어, 발달, 영향, 변인, 사회적, 고독적, 지원, 방안, 모색, 궁극적, 아동, 발달, 외국어, 학습, 한국사회의, 안정, 정착, 도모, 목적, 수행, 다음, 연구목표, 국제결혼, 가정, 아동, 언어능력, 대학, 수용언어, 표현, 언어, 발달정도, 국제결혼, 가정, 언어, 환경, 셋째, 아동, 언어발달, 어머니, 한국사회, 정착, 사회적, 교육적, 지원, 요구, 사항, 문제, 해결, 방안, 모색
	연구, 문제점, 착안, 기존, 다양, 문헌, 바탕, 기업, 사회적, 책임, 활동, 바람직, 반영, 개념, 기초, 측정, 척도, 개발, 목적, 척도



전체			1개 논문 2편 이상 출현		
NO	키워드	출현빈도	NO	키워드	출현빈도
1	이민자	2	1	이민자	2
1	물론	2	1	물론	2
1	여성	2	1	여성	2
1	이름	2	1	이름	2
1	각종	2	1	각종	2
1	가정	1	2	분석	2
1	대가족	1	2	선적	2
1	말살	1	7	희망	2
1	국적	1	7	부족	2
1	농촌	1	9	외국어	2
1	이민	1	9	연구	2
1	사회	1	9	학문적	2
1	도시	1	10	학습	3
1	통일	1	10	제시	2
1	분석	1	11	모세	2
1	제시	1	11	사회	2
1	사회통합	1	11	대안	2
1	중심	1	12	대학	2
1	연구	1	12	영어교육	2
1	자료	1	14	임대	3
1	조사	1	14	시제	3
1	논문	1	14	제시	2
1	원래지	1	23	서사	13
1	비유	1	15	환자	10
1	지역	1	15	문화적	5
1	지역사회	1	15	문화적	4
1	강령	1	15	강령	3
1	문화적	1	15	전개	2
1	지역공동체	1	15	변화	2

〈그림 1〉 문장의 형태소 분석 및 명사 형태 단어 추출 정리 작업



〈그림 3〉 논문 한 편 당 'Research Objectives'에 해당하는 내용의 어절 수의 분포

보면 '따라서'가 가장 많이 출현하였으며, '이에', '그리고', '또한', '즉' 순서로 나타났는데(〈표 3〉 참조), 출현한 위치를 기준으로 살펴보면 '따라서'와 '이에'가 162편의 논문 내에서 연구 목적을 기술하는 문단의 첫 어절에 위치하고 일부는 문단 내에 위치하였으며, '그리고', '또한'은 기술 문단 중간에서, '즉'은 1편의 논문에서 연구 목적 기술 문단의 첫 어절에 위치하고 나머지 논문들에서는 문단 내에 위치하는 것으로 분석되었다.

이와 같은 접속어의 성격 및 문장 내 위치는, 많은 저자들이 연구 목적의 내용을 기술함에 있어 이전에 해당 분야에서 수행되어 왔던 선행연구들의 경향이나 세부 내용 등을 언급한 후 기존 연구 결과들에 덧붙여 더 필요한 내용들을 본인 논문에서 어떻게 보완해 연구할 것 인지를 연결해 서술하는 과정에서 인과 관계를 나타내는 접속어를 많이 사용하기 때문인 것으로 파악된다.

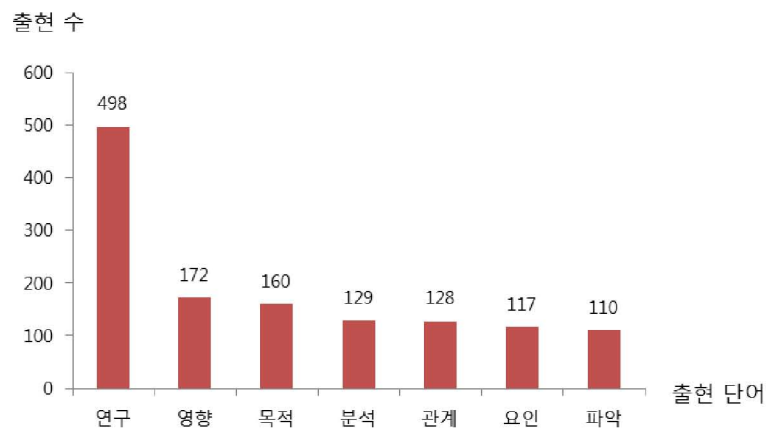
형태소 분석을 통해 추출한 명사 형태 단어들에 포함된 어절들을 분석한 결과, 전체 논문들에서 'Research Objectives' 항목을 기술하는데 있어 많이 사용되는 단어는 100회 이상 출현한 단어를 기준으로 파악했을 때 '연구(498회)', '영향(172회)', '목적(160회)', '분석(129회)', '관계(128회)', '요인(117회)', '파악(110회)'인 것으로 나타났다(〈그림 4〉 참조).

이 중 '연구'라는 단어는 전체 출현빈도 중 약 59%에 해당하는 293건의 횟수로 문장 내에서 주어의 역할로 사용되고 있었으며, '연구는(156회, 31%)', '연구에서는(137회, 28%)'과 같이 기술되었다. 그 외에 '연구의(80회, 16%)' 형태로 수식어 역할로도 사용되었는데 뒤에는 '목적'이라는 단어를 약 80%의 비율로 가장 많이 수식하고 있었고 그 외에 '일환', '필요성', '목표' 등의 단어를 수식하였다.

'영향'이라는 단어는 약 69%에 해당하는 118건의 횟수로 문장 내에서 '영향을' 형태로 목적

〈표 3〉 ‘Research Objectives’ 항목에 출현하는 접속어 출현 수

구분	분석 대상 접속어	출현 수	종류별 출현 수
순접	그리고	35	35
역접	그러나	2	3
	하지만	1	
	그럼에도(불구하고)	0	
인과	그래서	0	203
	따라서	112	
	이에	79	
	그러므로	9	
	때문에	3	
대등	또는	2	22
	또한	20	
첨가/보충	게다가	0	0
	왜냐하면	0	
	다시(말하면)	0	
	더구나	0	
전환	그런데	0	0
	단	0	
	다만	0	
	한편	0	
	오히려	0	
예시	가령	0	1
	예를(들면/들어)	1	
	이렇게면	0	
환언/요약	요컨대	0	19
	즉	16	
	결국	3	



〈그림 4〉 ‘Research Objectives’ 항목에 다수 출현한 단어 수

어 역할로 많이 사용되었고, '목적'이라는 단어는 약 79%에 해당하는 126건이 주어 역할로 사용되었고, 10%에 해당하는 16건이 목적어 역할로 사용되고 있었는데 주어 형태는 '목적은(67회, 42%)', '목적이(34회, 21%)', '목적으로(25회, 16%)' 순으로, 목적어 형태는 '목적을(16회, 10%)'로 많이 기술되었다.

'분석'이라는 단어는 약 71%에 해당하는 92건의 횟수로 문장 내에서 서술어 역할로 사용되었다. 어미 변화 및 다른 단어와의 합성에 따라 매우 다양한 형태로 출현하였는데 가장 많이 사용된 형태 순으로 정리하면 '분석하고자(17회, 13%)', '분석한다(10회, 8%)', '분석함으로써(9회, 7%)', '분석하여(8회, 6%)', '분석하고(6회, 5%)', '분석하며(6회, 5%)' 등으로 정리해 볼 수 있다.

'관계'라는 단어는 다른 단어와 합성되는 부분(인과관계, 상관관계 등)을 포함했을 때 약 53%에 해당하는 68건의 횟수로 문장 내에서 목적어 역할로 사용되었다. 목적어로 사용된 형태는 다른 단어와 합성되는 내용을 포함해 '관계를(50회, 39%)', '상관관계를(11회, 9%)', '인과관계를(7회, 5%)'와 같이 나타났다. 그 외에는 '있는지(를)'이라는 서술어와 함께 '관계가(10회, 8%)' 및 '상관관계가(3회, 3%)'의 주어 형태로, 또한 '대하여(대해)' 앞에 위치하며 '관계에(6회, 5%)'라는 형태 등으로 문장 내에서 사용되었다.

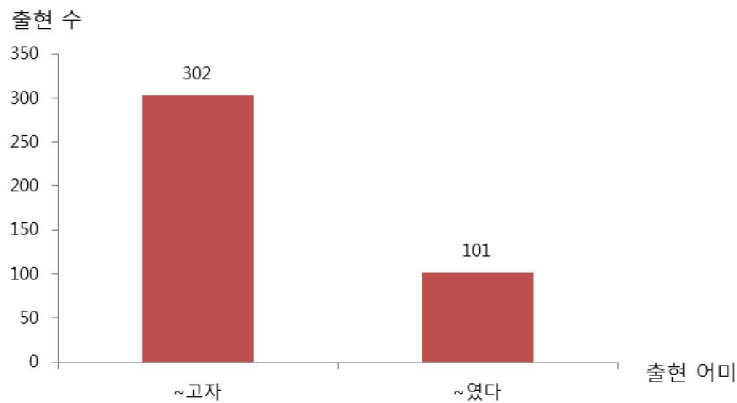
'요인'이라는 단어는 연구 목적의 세부 내용을 상술하는 데 있어 많이 사용된 단어로 특히 '규명하다/파악하다/확인하다'와 같은 종류의 서술어 앞에서 '요인을(19회, 16%)', '요인들을(9회, 8%)' 처럼 목적어 형태로 사용되거나 연

구에서 주목하는 여러 요인들이 다른 내용과 어떤 관계로 어떤 영향을 형성하는지에 대한 서술 문장 내에서 '요인이(9회, 8%)', '요인들이(5회, 4%)', '요인과(4회, 3%)'의 형태로 많이 나타났다.

'파악'이라는 단어는 약 94%에 해당하는 103건의 횟수로 문장 내에서 서술어 역할로 사용되었다. 전술한 '분석'이라는 단어와 마찬가지로 어미 변화에 따라 다양한 형태가 나타났는데, 많이 사용된 순으로 보자면 '파악한다(42회, 38%)', '파악하고(24회, 22%)', '파악하고자(10회, 9%)', '파악하여(8회, 7%)', '파악함으로써(6회, 6%)' 등으로 정리할 수 있다.

한편, 문장 내 동사의 어미 변화의 패턴을 가지고 분석한 결과, 'Research Objectives' 항목을 기술하는 데 사용되는 동사에 100회 이상 출현한 어미 형태는 '~고자(302회)', '~였다(101회)'로 나타났다(<그림 5> 참조).

'~고자'는 여러 어간 뒤에 붙어 어떤 행동을 할 의도나 욕구를 나타내는 의미의 연결 어미로 'Research Objectives' 항목을 기술하는 문장에서 가장 많이 출현하였다. '~고자' 앞에 같이 연결되는 내용을 살펴보면, '살펴보고자(33회, 11%)', '보고자(23회, 8%)', '제시하고자(17회, 6%)', '분석하고자(17회, 6%)', '제공하고자(16회, 5%)', '알아보고자(14회, 5%)', '규명하고자(14회, 5%)', '파악하고자(10회, 3%)' 순으로 나타났다. 이 중 '보고자'의 경우 앞부분에 '논의해/분석해/검증해/고찰해/규명해/살펴' 등의 동사를 서술하는 데 보조적인 역할을 수행하는 어미로 파악되어 의미 분석에 있어서는 앞에 연결되는 동사로 파악함이 적절할 것으로 판단된다.



〈그림 5〉 ‘Research Objectives’ 항목에 다수 출현한 어미 수

‘~였다’는 어떤 행위가 완료되는 상태를 나타내는 의미의 연결 어미로, ‘Research Objectives’ 항목에서 사용된 순서를 살펴보면 ‘하였다(54회, 53%)’, ‘수행하였다(6회, 6%)’, ‘조사하였다(5회, 5%)’, ‘검토하였다(5회, 5%)’, ‘설정하였다(4회, 4%)’, ‘분석하였다(3회, 3%)’ 순으로 나타났다. 이 중 ‘하였다’의 경우는 앞부분에 ‘~고자’에서 언급된 주요 동사들, 즉 ‘살펴보고자/제시하고자/분석하고자’ 등과 같이 연결되어 나타나는 경향을 띄어 앞서 기술한 ‘보고자’와 마찬가지로 앞에 연결되는 동사로 의미 분석을 하는 것이 적절한 것으로 분석되었다.

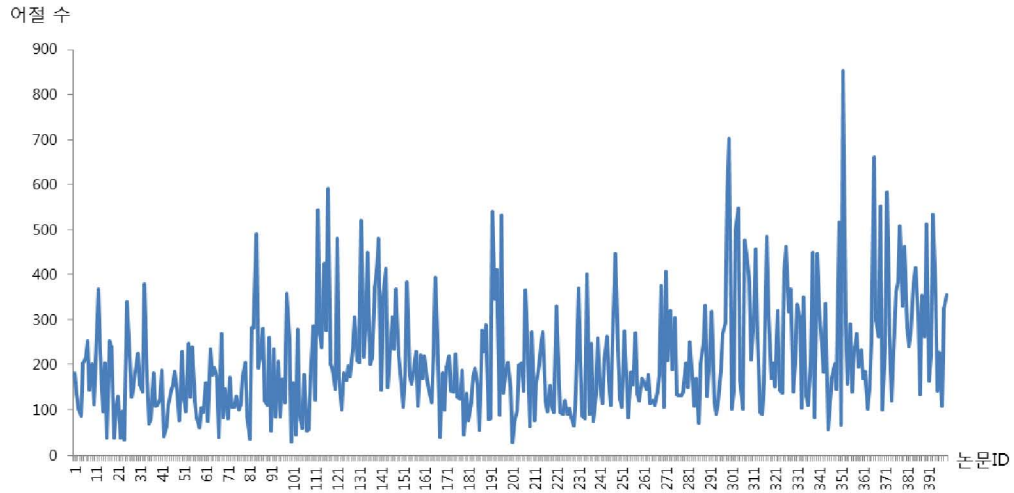
2.2 ‘Research Outcomes’ 항목의 문장 분석

Research Outcomes 항목에 입력된 콘텐츠의 경우, 전체 논문에서 수집된 데이터의 크기를 어절 단위로 분석해 본 결과 논문 한 편 당 최소 27개 어절부터 최대 852개 어절까지 분포되어 있었으며 평균 212개 어절로 나타나, ‘Research Objectives’에 비해 약 5.6배 가량 많은 어절로 구

성되어 있는 것으로 파악되었다(〈그림 6〉 참조).

‘Research Outcomes’ 항목의 기술을 위해 대상 논문들에서 주로 사용된 접속어 종류를 분석한 결과, 연구 목적 부분과 마찬가지로 인과 관계를 나타내는 접속어가 총 283회로 전체 논문의 결과 서술 내용에서 가장 많이 출현하였으며, 뒤를 이어 대등관계, 순접 관계, 환언/요약 관계를 나타내는 접속어 순으로 출현하였다. 하지만 실제 사용된 접속어 각각을 기준으로 보면 대등 관계에 속하는 ‘또한’이 가장 많이 출현하고, ‘그리고’, ‘즉’, ‘따라서’, ‘그러나’ 순서로 나타났다(〈표 4〉 참조). 출현 위치의 경우 ‘이에’가 2편의 논문 내에서 연구 결과를 기술하는 문장의 첫 어절에 위치한 것을 제외하고는, ‘그리고’, ‘즉’, ‘따라서’ 접속어 모두 문단 중간에 위치하는 것으로 분석되었다.

연구 결과 부분의 경우, 대체로 논문 중반부에서 분석한 연구 내용들의 단순 요약과 해당 내용에서 파악할 수 있는 함의에 대한 내용이 기술되기 때문에, 요약 내용을 같은 자격으로 나열하는데 사용되는 접속어인 ‘또한’과 앞의 내용을 이어 연결하는 ‘그리고’가 많이 사용되



〈그림 6〉 논문 한 편 당 'Research Outcomes'에 해당하는 내용의 어절 수 분포

〈표 4〉 'Research Outcomes' 항목에 출현하는 접속어 출현 수

구분	분석 대상 접속어	출현 수	종류별 출현 수
순접	그리고	188	188
역접	그러나	109	146
	하지만	33	
	그럼에도(불구하고)	4	
인과	그래서	2	283
	따라서	148	
	이에	61	
	그러므로	7	
	때문에	65	
대등	또는	48	271
	또한	223	
첨가/보충	게다가	1	47
	왜냐하면	2	
	다시(말하면)	42	
	더구나	2	
전환	그런데	7	89
	단	2	
	다만	2	
	한편	56	
	오히려	22	
예시	가령	1	6
	예를(들면/들어)	5	
	이렇게면	0	
환언/요약	요컨대	3	175
	즉	157	
	결국	15	

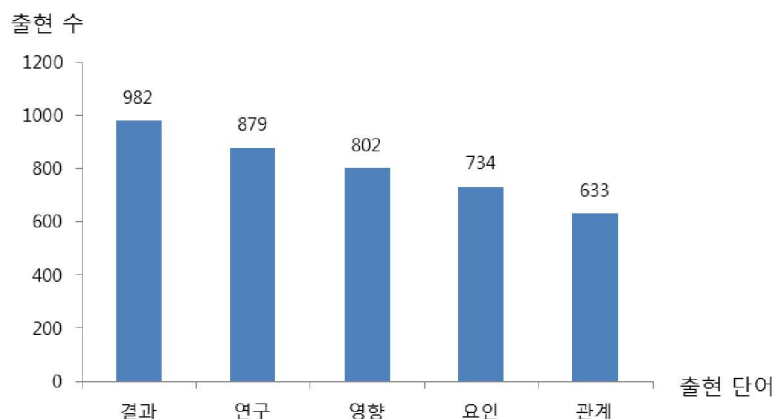
는 것으로 판단된다. 이어 연구 내용에 대한 부분을 쉽게 풀어 쓰거나 연구 진행의 결과에 대한 내용을 기술하는 데에는 환언/요약의 관계를 나타내는 ‘즉’과 앞의 문장에 대한 인과 관계를 표현하는 ‘따라서’가 함께 많이 쓰이는 것으로 여겨진다.

형태소 분석을 통해 추출한 명사 형태 단어가 포함된 항목들을 분석한 결과, 전체 논문들에서 ‘Research Outcomes’ 항목을 기술하는데 있어 많이 사용되는 단어는 500회 이상 출현한 단어를 기준으로 파악했을 때 ‘결과(982회)’, ‘연구(879회)’, ‘영향(802회)’, ‘요인(734회)’, ‘관계(633회)’인 것으로 나타났다(〈그림 7〉 참조).

이 중 ‘결과’라는 단어는 앞 부분에 다른 단어와 합성되는 내용(연구/분석/측정/조사 등), 즉 ‘결과(307회, 31%)’, ‘연구결과(21회, 2%)’, ‘분석결과(9회, 1%)’의 형태들을 포함해 보았을 때 전체의 37%에 해당하는 360건이 접속어의 역할로 사용되는 것으로 파악되었다. 이어서 약 22%에 해당하는 220건의 단어가 주어

역할로 사용되었는데, 주로 사용된 형태는 ‘결과는(158회, 16%)’, ‘결과가(22회, 2%)’, ‘연구 결과는(21회, 2%)’ 등으로 나타났다. 그 외에도 ‘결과를(102회, 10%)’, ‘연구결과를(19회, 2%)’과 같은 목적어 역할과 ‘결과이다(21회, 2%)’, ‘결과였다(6회, 0.1%)’와 같이 서술어 역할에서도 해당 단어가 사용되었다.

‘연구’라는 단어는 과반 이상이 주어 역할로 사용되었던 ‘Research Objectives’ 항목과 달리, ‘Research Outcomes’에서는 다른 단어를 수식하는 수식어의 역할 및 연구의 내용이 기술되는 장소를 나타내는 부사어의 역할과 주어의 역할이 비슷한 비율로 나타났다. 수식어 역할의 경우 ‘연구의(106회, 12%)’, ‘연구(71회, 1%)’의 형태로, 부사어의 경우 ‘연구에서(96회, 11%)’, ‘연구에서는(71회, 1%)’, ‘연구에(18회, 0.2%)’의 형태로, 주어의 경우 ‘연구는(69회, 1%)’, ‘연구가(37회, 0.4%)’ 등의 형태로 기술되었다. ‘연구’라는 단어의 수식을 받는 연결 단어로는 ‘(주요) 결과가 가장 많았고, 그 밖에 ‘변인’, ‘개념’, ‘과정’, ‘방법’ 등의 단어들이 함께 사용되었다.



〈그림 7〉 ‘Research Outcomes’ 항목에 다수 출현한 단어 수

‘영향’이라는 단어는 ‘영향력’이라는 단어와 함께 ‘영향을(581회, 72%)’, ‘영향력을(44회, 6%)’의 형태로 목적어 역할로 많이 기술되었고, 그 외에는 ‘영향력이(30회, 4%)’, ‘영향이(16회, 2%)’와 같이 주어 역할로 많이 사용되는 것으로 분석되었다.

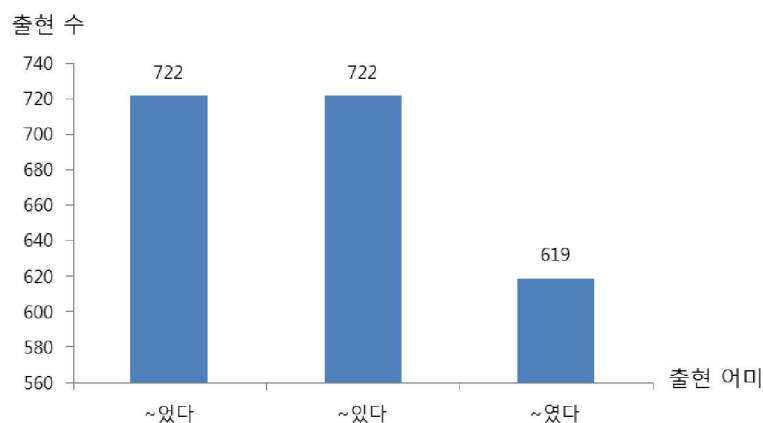
‘요인’이라는 단어는 ‘Research Objectives’ 항목에서 사용된 것과 마찬가지로 주어나 목적어와 같은 특정 역할로 많이 사용되기 보다는 연구가 진행된 내용을 설명하기 위한 목적으로 다양한 형태로 많이 사용되었다. 즉, ‘분석’이라는 단어와 합성되어 통계 분석 방식의 하나인 ‘요인 분석’ 형태(33회, 5%)로, 연구 결과가 나타나게 된 원인을 구체적으로 설명하기 위해 ‘요인에(19회, 3%)’, ‘요인과(19회, 3%)’, ‘요인의(15회, 2%)’, ‘하위요인(14회, 2%)’, ‘요인들이(10회, 1%)’ 등의 형태로 많이 기술되었다.

‘관계’라는 단어는 세부 출현 비율과 빈도에 있어서는 차이가 있지만 ‘Research Outcomes’에서도 ‘Research Objectives’와 유사한 역할로 많이 사용되었다. ‘상관관계’, ‘정적관계’, ‘부

적관계’, ‘인과관계’ 등과 같이 다른 단어와 함께 사용되는 내용을 포함해 약 23%에 해당하는 147건의 횡수로 문장 내에서 목적어 역할로 사용되었다. 목적어로 사용된 형태는 다른 단어와 합성되는 내용을 포함해 ‘관계를(82회, 13%)’, ‘상관관계를(31회, 5%)’, ‘정적관계를(9회, 2%)’, ‘부적관계를(4회, 1%)’과 같이 나타났다. 그 외에는 ‘있다’, ‘있었다’, ‘없었다’와 같은 서술어 앞에서 ‘관계가(61회, 10%)’ 및 ‘상관관계가(18회, 3%)’의 주어 형태, 연구 내용의 세부 설명을 위한 ‘관계(27회, 4%)’, ‘관계에서(21회, 3%)’, ‘관계에(20회, 3%)’ 형태 등으로 문장 내에서 사용되었다.

문장 내 동사의 어미 변화의 패턴을 가지고 분석한 결과, ‘Research Outcomes’ 항목을 기술하는 데 사용되는 동사에 500회 이상 출현한 어미 형태는 ‘~었다(722회)’, ‘~ 있다(722회)’, ‘~였다(619회)’로 나타났다(〈그림 8〉 참조).

‘~었다’는 어떤 행위의 완료를 표현해주는 의미로 사용되는데 ‘Research Outcomes’ 항목에서 가장 많이 출현한 형태는 ‘있었다(235회,



〈그림 8〉 ‘Research Outcomes’ 항목에 다수 출현한 어미 수

33%)'였으며, 주로 '~(알/할/불) 수', '~경향이' 등과 연결되어 사용되었다. 그 외에는 연구의 내용을 드러내거나 정리해주는 의미의 '나타내었다(67회, 9%)', '없었다(40회, 6%)', '확인되었다(33회, 5%)', '분석되었다(15회, 2%)', '보여주었다(14회, 2%)' 등의 형태로 기술되고 있었다.

'~있다'는 존재의 상태를 표현해주는 의미로 'Research Outcomes'에서 가장 많이 출현한 형태는 '~(알/할/불) 수'와 연결되어 사용되는 '있다(591회, 77%)'로 나타났다. 해당 동사는 앞서 설명한 '있었다'와 유사한 방식으로 사용되는 것으로 파악되며, 각각 논문의 결과 부분에서 연구된 내용들을 확인하는 과거형과 연구에서 내린 결론의 의미 등을 해석하는 현재형을 구분하는 데 사용되는 것으로 볼 수 있다. 그 외에는 문장 연결을 위한 조사와 함께 '있다는(73회, 10%)', '있다고(53회, 7%)'의 형태로 사용되고 있었다.

'~였다'는 '~있다'와 마찬가지로 행위의 완료를 나타내는 의미로 사용되지만 앞에 함께 오는 단어의 받침 유무에 따라 달리 사용되는 어미이다. 'Research Outcomes' 항목에서는 '보였다(112회, 18%)', '하였다(45회, 7%)', '증가하였다(35회, 6%)', '분석하였다(21회, 3%)', '도출하였다(16회, 3%)', '측정하였다(14회, 2%)'와 같이 사용되었는데, 이 중 '하였다'는 'Research Objectives'에서 사용된 방식과 마찬가지로 '살펴보고자/무색하고자/알아보고자' 등의 동사와 많이 연결되는 경향을 보였고 그 외에는 '시도를/비교를/검토를' 등의 목적어와 함께 '하다'의 과거형으로 사용되었다.

3. 결 론

본 연구는 연구논문의 논리적 의미 구조 메타데이터 항목에 해당하는 콘텐츠에 담겨 있는 문장의 구성에 따라 시스템에서 적절한 항목으로 자동 식별·처리될 수 있도록 하는, 문장의 미론(sentence semantics)적 분석 방법을 제안하고자 하는 목적으로 수행되었으며, 의미 구조 메타데이터 항목 중 'Research Objectives'와 'Research Outcomes'에 해당하는 연구 논문 문장의 구조 분석을 통해 도출된 결과는 다음과 같다.

첫째, 'Research Objectives'와 'Research Outcomes'에 입력한 문장들의 어절 수를 분석한 결과 연구 목적에 해당하는 문장의 어절 수는 최소 4개부터 최대 141개(평균 38개), 연구 결과에 해당하는 문장의 어절 수는 최소 27개부터 최대 852개(평균 212개)까지 분포되는 것으로 분석되어 약 1 대 5.6의 비율을 나타냈다. 이는 연구의 결론 부분을 기술함에 있어 연구 목적 부분에 비해 논문에서 분석한 세부 결과들을 나열하고 요약하는 문장이 많이 삽입되기 때문인 것으로 여겨진다.

둘째, 접속어의 경우 'Research Objectives'는 인과-순접-대등-환언/요약 관계를 나타내는 접속어 순으로, 'Research Outcomes'는 인과-대등-순접-환언/요약 관계를 나타내는 접속어 순으로 많이 출현한 것으로 파악되었다. 구체적으로 논문에서 실제 사용된 접속어를 살펴보면 연구 목적을 기술하는 부분에서는 '따라서(인과)' - '이에(인과)' - '또한(대등)' - '즉(환언/요약)'이, 연구 결과 부분에서는 '또한(대등)' - '그리고(순접)' - '즉(환언/요약)' - '따라서

(인과)' 순으로 사용되었다. 또한 접속어의 출현 위치를 보았을 때 연구 목적에서 많이 사용되는 '따라서'나 '이에'와 같은 접속어는 문장의 첫 시작 부분에 많이 나타났고, 연구 결과에서 사용되는 접속어들은 주로 문단 내에서 문장들을 연결하는 부분에서 많이 나타났다. 이는 'Research Objectives'의 경우 많은 연구자들이 논문에서 다루는 주제와 관련한 기존 연구 경향 등을 서술하고 나서 그 내용에 기초해 본인의 연구가 어떻게 다르고 추가적으로 대비되는 부분이 무엇인지 강조하며 연구 목적을 기술하는 방식으로 서술하고, 'Research Outcomes'는 논문의 본문에서 다룬 연구 내용들을 요약해 나열한 후, 해당 내용에서 도출할 수 있는 주요 시사점들을 나열하는 방식으로 기술되는 논문 기술 방식의 특징 때문인 것으로 파악된다.

셋째, 형태소 분석을 통해 추출한 명사 형태 단어들 중 특정 주제 분야의 개념을 포함하는 키워드 등을 제외하고 일반적으로 논문의 내용을 기술하는 데 주로 사용되는 단어들을 중심으로 출현빈도 순으로 나열해보면 'Research Objectives'는 '연구' - '영향' - '목적' - '분석' - '관계' - '요인' - '파악'으로, 'Research Outcomes'는 '결과' - '연구' - '영향' - '요인' - '관계'로 정리해 볼 수 있었다. 각각의 단어들은 항목별로 문장 내에서 주어, 목적어, 서술어 역할 등으로 사용되고 있었으며, '역할'이나 '요인', '관계'는 목적어나 결과 부분 모두에서 비슷한 역할을 담당하고 있었지만 '연구'는 같은 단어라도 목적 부분과 결과 부분에서 사용되는 역할에 차이를 보였다. 사실 이러한 단어들은 색인 작업을 할 때에는 지나치게 일반적인 단어로 분류되기 때문에 출현 빈도가 높더라도 유의한 색

인어로서의 역할은 하지 못한다. 하지만 논문의 의미 구조에 기반한 메타데이터 항목에 해당하는 문장들을 추출하는데 있어서는 논문 서술에 있어 문장 내에서 많이 사용되는 단어들이기 때문에 오히려 문장의 어떤 부분에서 어떤 역할로 사용되는지에 대한 규칙들을 도출할 수 있다면, 관련 항목들의 내용을 자동 추출하는 데에는 도움을 줄 수 있다고 판단된다.

넷째, 문장 내 동사의 어미 변화 패턴과 관련해서는 'Research Objectives'에서 많이 출현한 어미는 '~고자'와 '~였다', 'Research Outcomes'에서는 '~었다', '~있다', '~였다'로 나타났다. 동사에 적용되는 어미의 경우는 종류에 따라 어떤 의도나 상태 등을 나타내 주기는 하지만 앞부분의 어간에 붙는 내용에 따라 주요 의미가 달라지는데, 'Research Objectives'에서 많이 사용된 동사는 '살펴보고자', '보고자', '제시하고자', '하였다', '수행하였다', '조사하였다' 등이었고, 'Research Outcomes'에서 많이 사용된 동사는 '있었다', '나타내었다', '없었다', '있다', '있다는', '있다고', '보였다', '하였다', '증가하였다' 등이었다. 이 중 '하였다'는 목적 부분과 결론 부분 모두에서 '~고자' 형태의 어미가 붙은 동사들과 연결되어 많이 사용되고 있었으며, 결론 부분에서 많이 사용되는 '있다'와 '있었다'는 '~(알/할/볼) 수' 등과 같이 처리 방법이나 능력을 나타내는 단어와 주로 함께 사용되었다.

본 연구는 KCI에서 제공되는 전체 학술논문 데이터 중에서 피인용 빈도가 높은 일부 논문들만을 분석 대상으로 하고 있다는 점, 그리고 분석 대상인 모든 논문에서 해당 내용이 입력된 '연구 목적(Research Objectives)'과 '연구

결과(Research Outcomes)’에 대한 메타데이터 항목의 데이터에 대해서만 분석이 이루어졌다는 점에서 전 분야의 모든 논문들의 관련 내용에 일반화시키기에는 한계가 있다고 할 수 있다. 또한 분석 대상 단어의 기준을 정함에 있어, 각 메타데이터 항목에 기술된 문장들의 어절 수 평균에 비례하여 출현 횟수 기준을 달리해 분석을 수행하였다는 점도 어느 정도 한계를 지닌다고 할 수 있겠다. 하지만 본 연구의 데이터에 대한 분석 방식은 다른 유사 연구들에서 수행된 경우가 거의 없기 때문에 관련 데이터의 분석 기준을 어느 수준으로 정할 것인가에 대한 참고 자료도 미비하다고 할 수 있다. 때문에 이러한 한계점들은 향후 분석 대상 데이터의 범위를 다양화하고, 출현빈도나 위치, 품사 외에도 학문 분야나 주제, 분석 대상 텍스트의 장르 등 여러 가지 분석 기준을 세분화해 비교·적용해 보는 후속 연구들을 통해 보완이 가능할 것으로 판단된다.

이상의 한계점에도 불구하고, 본 연구는 연구자의 학술적 이해형성을 지원하기 위해 연구 논문이 담고 있는 공통된 논리적 의미를 반영한 메타데이터 요소의 자동 식별과 입력 방안을 제시하는 데 활용할 수 있는 기초 연구로서 의의가 있다고 할 수 있다. 자연어 처리 분야에서 일반 문서의 자동 요약 처리를 위한 한국어 문장의 자동 추출 방식은 아직까지 대상 문서 내에 포함된 문장의 중요도를 문장 길이나 출현 단어 수 등을 기반으로 가중치 등을 계산해 조합·추출해내는 방식이기 때문에 사람과 같은 수준의 작업은 어렵다. 하지만 서론에서 언

급했듯이 연구 논문은 대체로 각 내용을 구성하는 논리적 의미 구조가 체계화 되어 있기 때문에, 본 연구에서 분석한 방법을 토대로 논문 작성 시 연구자들이 연구 목적, 방법, 결과 등의 항목을 서술할 때 주로 사용하는 문장 구문의 구성 요소들(많이 사용되는 접속어, 주어와 서술어, 문장 내 위치 등)과 표현 방식(자주 사용되는 단어의 표현, 어미 변화 패턴 등) 등을 데이터로 누적해 각 항목에 해당하는 문장 추출 알고리즘을 구성하는 기반 데이터로 활용하면, 관련 시스템을 구축하고자 할 때 사람이 직접 논문을 읽고 일일이 찾아 입력하지 않더라도 논문 내 연구목적(따라서 본 연구의 목적은 ~ 것이다)이나 결과(연구 결과 ~ 확인되었다) 등을 기술한 문장들을 추출해 DB화하는 작업이 가능할 것이다.

아울러 본 연구에서 수행된 내용은 한국어로 작성된 빅데이터 처리 기술 분야의 비정형 데이터 분석에도 응용이 가능할 것으로 사료되며, 궁극적으로 문장의미론에 기반한 한국어 의미 처리 및 검색 기법 분야에서의 기술적인 향상에도 조금이나마 기여할 수 있을 것으로 기대된다. 또한 본 연구를 토대로 후속 연구들이 이어져 각 분야의 연구 논문에 포함된 논리적 지식 구조를 구성하는 문장 특징을 분석해 관련 시스템을 구축하는 데 활용할 수 있게 된다면, 연구자들이 연구를 수행하는 데 있어 유사 연구 성과들을 분석하는 데 도움을 줄 수 있는 의미적 검색 시스템의 DB 구축 과정에 소요되는 비용 절감과 함께 입력 데이터의 질적 수준에 대한 일관된 품질을 보장할 수 있을 것이다.

참 고 문 헌

- 강범일, 송민, 조화순 (2013). 토픽 모델링을 이용한 신문 자료의 오피니언 마이닝에 대한 연구. 한국문헌정보학회지, 47(4), 315-334. <http://doi.org/10.4275/KSLIS.2013.47.4.315>
- 고영만, 송인석 (2011). 연구문헌의 지식구조를 반영하는 의미기반의 지식조직체계에 관한 연구. 정보관리학회지, 28(1), 145-170. <http://doi.org/10.3743/KOSIM.2011.28.1.145>
- 김민철, 심규승, 한남기, 김예은, 송민 (2013). 트위터 상의 악의적 이용 자동분류. 한국문헌정보학회지, 47(1), 269-286. <http://doi.org/10.4275/KSLIS.2013.47.1.269>
- 김진옥, 이선숙, 용환승 (2011). 한글 텍스트의 오피니언 분류 자동화 기법. 정보과학회논문지: 데이터베이스, 38(6), 423-428.
- 미래창조과학부 (2015). 과학기술논문(SCI) 분석 연구. 연구보고서.
- 송민선, 고영만 (2015). 한국학 연구 논문의 의미구조 기반 메타데이터 연구. 한국도서관·정보학회지, 46(3), 277-299. <http://doi.org/10.16981/kliiss.46.3.201509.277>
- 송민선, 고영만, 이승준 (2016). 한국학 연구 논문의 텍스트 구조 기반 메타데이터 검색 시스템 개발 연구. 정보관리학회지, 33(3), 155-176. <http://doi.org/10.3743/KOSIM.2016.33.3.155>
- 신준철, 옥철영 (2016). 한국어 의미분석을 위한 의미자원과 어휘의미분별. 정보과학회지, 34(8), 8-16
- 안애립, 한용진, 박세영, 남지순 (2012). 휴먼 오피니언 자동 분류 시스템 구현을 위한 비결정 오피니언 형용사 구문 처리. 정보과학회논문지: 컴퓨팅의 실제 및 레터, 18(2), 158-162.
- 유사라, 이혜원, 송인석 (2009). 연구자 중심 연구성과물 의미검색을 위한 인문사회 학술용어 온톨로지 적용 및 유지관리 체계 연구. 한국문헌정보학회지, 43(2), 277-298.
- 윤구호 (1999). 자동색인의 이론과 실제. 한국도서관·정보학회지, 30(3), 27-51.
- 윤성희, 백선욱 (2004). 단어 의미 정보를 활용하는 이용자 자연어 질의 유형의 효율적 분류. 정보관리학회지, 21(4), 251-263.
- 정영미, 이태영 (1982). 자동색인(自動索引)의 통계적기법(統計的技法)과 한국어문헌(韓國語文獻)의 실험(實驗). 도서관학, 9, 99-118.
- 한국학술지인용색인. Retrieved from <http://www.kci.go.kr>
- 한정기, 박민규, 김준태 (1998). 구문 패턴과 키워드 집합을 이용한 자동 문서 분류의 성능 향상. 한국정보과학회 인간과 컴퓨터 상호 작용 연구회 학술 대회 발표 논문집(HCI), 70-73.
- Harmsze, F. A. P. (2000). A modular structure for scientific articles in an electronic environment. Ph. D. Dissertation, Vander Waals-Zeeman Institute, University of Amsterdam.
- Horn, R. (2000). Teaching philosophy with argumentation maps. Newsletter of the American Philosophical Association, November, 2000.

- Kando, N. (1997). Text-level structure of research articles and its implication for text-based information processing systems, *Proceedings of the 19th British Computer Society Annual Colloquium on Information Retrieval Research*, Aberdeen, Scotland, UK, April, 1997, 68-81.
- Kando, N. (1999). Text structure analysis as a tool to make retrieved documents usable, *Proceedings of the 4th International Workshop on Information Retrieval with Asian Languages*, Taipei, Taiwan, November 11-12, 126-135.
- The Royal Society (2011). Knowledge, networks, and nations: global scientific collaboration in the 21st century. RS Policy Document 03/11 Retrived from https://royalsociety.org/~media/Royal_Society_Content/policy/publications/2011/4294976134.pdf

• 국문 참고문헌에 대한 영문 표기 (English translation of references written in Korean)
--

- Ahn, Ae-Lim, Han, Yong-Jin, Park, Se-Young, & Nam, Jee-Sun (2012). Processing of adjectives of non-deterministic opinion for human opinion classification system, *Journal of KIISE: Transactions on Computing Practices*, 18(2), 158-162.
- Han, Junggee, Park, Minkyu, & Kim, Juntae (1998). Improving the performance of automatic text categorization by using phrasal patterns and keyword sets, *Proceedings of KISS Conference (HCI)*, 70-73.
- Jung, Youngmi, & Lee, Tae-Young (1982). Statistical techniques for automatic indexing and some experiments with Korean documents, *Library Science*, 9, 99-118
- Kang, Beomil, Song, Min, & Jho, Whasun (2013). A study on opinion mining of newspaper texts based on topic modeling, *Journal of the Korean Society for Library and Information Science*, 47(4), 315-334, <http://doi.org/10.4275/KSLIS.2013.47.4.315>
- Kim, Jin-Ok, Lee, Sun-Sook, & Yong, Hwan-Seung (2011). Automatic classification scheme of opinions written in Korean, *Journal of KIISE: Database*, 38(6), 423-428.
- Kim, Meen Chul, Shim, Kyu Seung, Han, Nam Gi, Kim, Ye Eun, & Song, Min (2013). Automatic classification of malicious usage on Twitter, *Journal of the Korean Society for Library and Information Science*, 47(1), 269-286, <http://doi.org/10.4275/KSLIS.2013.47.1.269>
- Ko, Young-Man, & Song, Inseok (2011). A study on the knowledge organizing system of research papers based on semantic relation of the knowledge structure, *Journal of the Korean Society for Information Management*, 28(1), 145-170, <http://doi.org/10.3743/KOSIM.2011.28.1.145>

- Korea Citation Index. Retrieved from <http://www.kci.go.kr>
- Ministry of Science, ICT and Future Planning (2015). SCI Analysis Research.
- Shin, Joon-Choul, & Ock, Cheol-Young (2016). Semantic resources for Korean semantic analysis and word sense disambiguation. *Journal of KIISE*, 34(8), 8-16
- Song, Min-Sun, & Ko, Young Man (2015). A study on the metadata based on the semantic structure of the Korean studies research articles. *Journal of Korean Library and Information Science Society*, 46(3), 277-299. <http://doi.org/10.16981/kliss.46.3.201509.277>
- Song, Min-Sun, Ko, Young Man, & Lee, Seung-Jun (2016). A study on developing a metadata search system based on the text structure of Korean studies research articles. *Journal of the Korean Society for Information Management*, 33(3), 155-176. <http://doi.org/10.3743/KOSIM.2016.33.3.155>
- Yoo, Sa-Rah, Lee, Hye-Won, & Song, Inseok (2009). A study on the application and management framework of social science scholarly ontology for semantic information navigation. *Journal of the Korean Society for Library and Information Science*, 43(2), 277-298.
- Yoon, Ku-Ho (1999). Theory and practice of automatic indexing. *Journal of the Korean Library and Information Science Society*, 30(3), 27-51.
- Yoon, Sung-Hee, & Paek, Seon-Uck (2004). Efficient classification of user's natural language question types using word semantic information. *Journal of the Korean Society for Information Management*, 21(4), 251-263.