제28회 한글 및 한국어 정보처리 학술대회 논문집 (2016년)

한국어 수사구조 분류체계 수립 및 주석 코퍼스 구축

노은정[O,†], 이연수[†], 김연우[††], 이도길[†††]
(주)엔씨소프트[†], 고려대학교 언어학과[††], 고려대학교 민족문화연구원[†††]
{enoh, yeonsoo}@ncsoft.com, wiskingdom@gmail.com, motdg@korea.ac.kr

Building an RST-tagged Corpus and its Classification Scheme for Korean News Texts

Eunchung Noh[O,†], Yeonsoo Lee[†], YeonWoo Kim[††], Do-Gil Lee[†††]
NCSOFT Corp.[O,†]; Department of Linguistics, Korea University[††], Research Institute of Korean Studies[†††]

요 약

수사구조는 텍스트의 각 구성 성분이 맺고 있는 관계를 의미하며, 필자의 의도는 논리적인 구조를 통해서 독자에게 더 잘 전달될 수 있다. 따라서 독자의 인지적 효과를 극대화할 수 있도록 수사구조를 고려하여 단락과 문장 구조를 구성하는 것이 필요하다. 그럼에도 불구하고 지금까지 수사구조에 기초한 한국어 분류체계를 만들거나 주석 코퍼스를 설계하려는 시도가 없었다. 본 연구에서는 기존 수사구조 이론을 기반으로, 한국어 보도문 형식에 적합한 30개 유형의 분류체계를 정제하고 최소 담화 단위별로 태깅한 코퍼스를 구축하였다. 또한 구축한 코퍼스를 토대로 중심문장을 비롯한 문장 구조의 특징과 분포 비율, 신문기사의 장르적 특성 등을 살펴봄으로써 텍스트에서 응집성의 실현 양상과 구문상의 특징을 확인하였다. 본 연구는 한국어 담화 구문에 적합한 수사구조 분류체계를 설계하고 이를 이용한 주석 코퍼스를 최초로 구축하였다는 점에서 의의를 갖는다.

주제어: 수사구조이론(Rhetorical Structure Theory), 주석 코퍼스(Tagged/Annotated Corpus), 담화 구조(Discourse structure), 코퍼스 분석(Corpus analysis)

1. 서론

본 연구는 한국어 수사구조에 기초한 담화 주석 코퍼스 구축을 목적으로 한다. 실제 언어 현상에서 문장의 의미와 기능이 텍스트의 관계 구조나 형식과 밀접하게 연관되어 실현된다는 점에서 해당 코퍼스의 구축과 확보는 학술적 목적은 물론 실용적 측면에서도 중요하다.

수사구조이론(Rhetorical Structure Theory, 이하 RST)은 William C. Mann과 Sandra A. Thompson이 제창한 텍스트 구조 기술을 위한 프레임으로서, 텍스트의 각 부분이 맺는 관계 방식과 종류를 토대로 담화 구조와 텍스트를 분석하는 이론이다[1]. 표면적 연결 관계를 의미하는 응결성(cohesion)과 의미상의 연결 관계를 일컫는 응집성(coherence)은 좋은 글이 가져야할 필수 속성으로서, 문장과 문단의 수사 구조와 밀접한 관련이 있다. 특히 수사구조이론의 응집성은 텍스트 내의 단어와 문장의 연쇄적인 상호 관계와 어휘 혹은 문법적인 의존 관계를 명시적으로 보여줄 수 있으며, 자연어 생성 시스템의 결과의 질을 평가하는 데에도 유용하다[2,3].

(1) a. 옷이 젖었다. 비가 내렸다.

b. ??옷이 젖었다. 철수는 음악 듣기를 좋아한다.

(1a)은 옷이 젖은 것이 비가 내렸기 때문이라는 인과적 설명이 가능하다. 반면에 (1b)의 음악 듣기를 좋아한다는 정보는 문장 간 응집성이 떨어진다고 볼 수 있다.

신문 기사 텍스트는 자연어 중에서도 정보성과 결속성이 높은 양질의 코퍼스라 볼 수 있다. 신문 기사는 응결성을 유지하기 위하여 문장과 단락이 서로 유기적으로 배열되고 선후 긴밀한 관계를 맺도록 연결되어야 한다. 또한 응집성을 위하여 일정한 유형의 구조나 어휘를 사용하거나 단락 전체 혹은 단락과 단락이 결합된 논리적인 의미 구조를 갖추는 데에 유념하여야 한다.

최근 로봇 저널리즘 등 자동 텍스트 생성에 관한 연구가 많아졌으나, 각 문장을 유기적으로 연결하기 위해 어떤 구조를 가져야하는가에 관한 연구는 일천하다. 해외에서 RST는 자동 텍스트 생산을 도출하기 위한 텍스트의 응집성 모델 연구로도 중요하게 개발되고 있으나, 현재까지 한국어 텍스트를 분석한 연구는 극히 드물다.

담화 분석과 텍스트 구조 분석을 위하여 실제 수사적 관계를 주석한 코퍼스에 대한 필요성은 계속 대두되고 있다. 텍스트는 일반적으로 그것이 사용되는 일정한 환

경, 즉 사용역(register)에 의해 구별된다. 언어가 실질적으로 어떻게 사용되는지 실제 언어 수행을 반영하는 코퍼스는 여러모로 효용 가치가 크다. 특히 구체적인 텍스트 조직 관계에 대한 연구는 기존의 코퍼스 연구에서는 파악할 수 없었던 실제 언어 사용의 상황에서 나타나는 다양한 정보를 기술하고 설명할 수 있다.

본 연구에서는 텍스트의 응결성과 응집성을 파악하고 기계에 학습시킬 데이터로서 기사문이 가장 적합하다고 판단하여, 이를 대상으로 수사구조를 분석하고 이를 주석한 코퍼스를 최초로 구축하고자 하였다.

2. 관련 연구

기존 한국어 코퍼스 연구는 실제 텍스트가 조직되는 양상에 대한 연구보다는 형태소 분석 내지는 문장 절 이내 분석에 국한되어 왔다[4]. 대개 한국어 학습자 대상 코퍼스나 특정 어휘의 빈도 혹은 공기어 분석, 번역문, 감성 어휘 분석 등에 한정된 경향이 있다.

한편 장르적 연구로서 신문 기사에 관한 선행연구에서는 주로 스타일과 작성법에 관한 연구 혹은 특정 기사 내용에 대한 비교 분석 등을 다루어 왔다[5]. 그리고 최근 텍스트 언어학적 관점에서 신문 기사문에 대한 텍스트성을 살펴보는 연구들이 등장하고 있다.

RST는 문맥의 앞과 뒤에 대한 구체적인 기준을 통해 의미 관계 분석이 가능하며, 심층 구조도 파악할 수 있다는 장점이 있다[6]. 하지만 지금까지는 주로 미시구조의 질적 분석을 위해 사용되었다. 선행연구는 주제별로 번역학적 관점에서 원천 언어와 목표 언어 사이에 발생하는 응집성 문제를 설명하거나[7], 텍스트언어학적 관점에서 특정 장르의 텍스트 유형에 따른 텍스트 구조를 밝히는 연구[3,8,9,10,11,12], 교육적 관점에서 국어 혹은 한국어 교육 효과를 높이고자 수사구조이론을 적용한 연구[13], 화용론이나 담화분석의 관점에서 생산자의 의도를 밝히고자 한 연구[14] 등에서 제한적으로 이루어졌다. 이후 신문 기사에 RST를 적용하여 텍스트분석을 시도한 연구[15]가 있었으나 6개 단락 총 15문장으로 이루어진 단일 기사에 대해서만 분석하였고, RST 전용 분석 툴을 사용하지 않고 관계 연결 과정을 설명하는 데에 그쳤다. 본 연구와 같이 분석툴을 활용하여 대량의 한국어 문서를 분석하고 수사구조이론을 코퍼스 구축과 데이터 분석에 활용한 사례는 없었다.

3. 분류 체계

수사구조이론은 관계(relations)와 도식(schemas)으로 구성된다. 수사 관계는 핵(nucleus)과 위성(satellite)이라 불리는 두 텍스트 단위에 적용되는 기능적 개념으로서 초기 Mann & Thompson (1988)은 23개의 목록으로 제시하였고 이후 학자들에 의해 개선을 거쳤다. 핵과 위성이 맺는 관계는 화살표와 선을 이용하여 도식화된다.

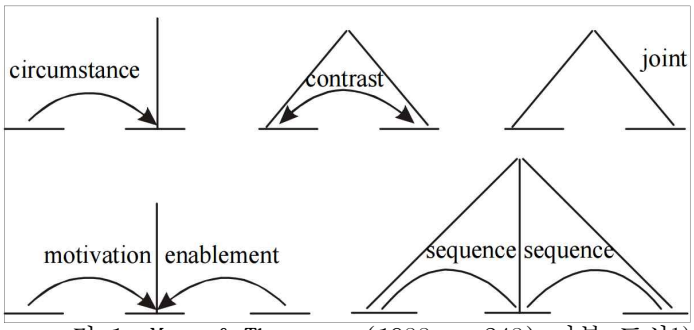

그림 1. Mann & Thompson (1988, p.248) 기본 도식[1)]

[그림 1]은 맺을 수 있는 기본 관계 구조를 나타낸다. 가령 대조(contrast)는 두 개 핵(N+N)으로만 구성되며, 유사점과 비교가능한 차이점을 독자에게 인식시키려는 의도를 갖고 있다. 두 N은 유사점을 갖되, 하나 혹은 몇 가지 부분에서 특징적인 차이점을 갖고 있어야 한다.

(2) [contrast] 구조 기사문 예시

N: LG는 4회에 3점을 뽑아내는 집중력을 보였다.

N: 반면 두산은 6회 7회 9회 한 점씩을 뽑아내며 끈질김을 보여줬다.

영어를 기초로 한 선행 이론은 기본적으로 구 또는 절 단위를 기초로 분석한다. 영어는 언어적 특성상 절 단위 분석이 가능하다. 절 단위로 분할할 경우 정교한 분석이 가능하고 문장 안의 수사 구조 분석이 가능하지만 유닛 수가 상대적으로 많아서 텍스트 분할 난이도는 물론 관계 분석 난이도 역시 높아진다. 이를 보완하여 RST의 최소 담화 단위(elementary discourse unit, 이하 EDU)를 설정하는 기준으로 언어적 형식(form 혹은 category), 기능(function), 의미(meaning)의 3요소가 거론되었다[16, 17]. 형식뿐만 아니라 통사적 요소의 기능과 명제 간 응집 관계를 복합적으로 고려하여 EDU를 설정하는 것이 바람직하다. 영어 외 언어들에서 이런 구분을 발견할 수 있다[18].

본 연구에서는 절 단위가 아닌 문장 단위로 스팬(span)을 구분한다. 언어 특성상 한국어의 EDU는 형식에만 기초한 분절이 어렵고 의미와 기능을 모두 고려하여 구분하는 것이 바람직하다. EDU를 문장으로 나눌 경우 종결어미와 문장부호를 통해 쉽게 구분할 수 있고 자동처리가 용이하다. 비록 분석 수준이 낮고 문장 안의 수사 구조(일부 조건문이나 상황 부사절 구문 등)에 대해서는 분석이 불가능하나, 스팬 간의 관계 연결에 더 집중할 수 있다.

선행 연구와 이론을 바탕으로 한국어 기사문에 적합하도록 수정을 거쳐 완성한 분류 기준은 [표 1,2]와 같다.

---

1) 출처 http://www.sfu.ca/rst/pdfs/RST_Introduction.pdf

표 1. 수사구조이론 분류

기존 수사구조 분류 (연구마다 목록에 차이가 있음)		NC-RIKS (2016)	
제시 관계	10개 유형 antithesis(반론), background(배경), concession(양보), enablement(가능화), evidence(증거), justify(정당화), motivation(동기화), preparation(도입), restatement(재진술), summary(요약)	**제시 관계**	동일
주제 관계	15개 유형 circumstance(환경), condition(조건), elaboration(정교화), evaluation(평가), interpretation(해석), means(수단), non-volitional cause(비의도적 원인), non-volitional result (비의도적 결과), otherwise(양자택일), purpose(목적), solutionhood(해결성), unconditional(무조건), unless(배제적 조건), volitional cause(의도적 원인), volitonal result(의도적 결과)	**주제 관계**	동일
다중 핵 관계	7개 유형 joint(연결) contrast(대조) sequence(연속) multinuclear restatement(다핵재진술) conjunction(접속) list(나열) disjunction(분리)	**다중 핵 관계**	4개 유형 joint(연결) contrast(대조) sequence(연속) multinuclear restatement(다핵재진술)
특수 기능 관계	2개 유형 same-unit(동일단위) quotation(인용)	**특수 기능 관계**	없음
기타 관계	없음	**기타 관계**	1개 유형 unstated-relation(기타관계)

특수 기능 관계는 절의 구성성분이라는 점을 전제로 하므로, 본 연구에서는 EDU를 문장 단위로 설정하면서 제외시켰다. 그리고 한국어 일부 기사문에서 말미에 전체 주제를 벗어나지는 않으나, 의미적으로 기존의 관계와는 다른 관계를 맺고 있는 단락이 등장하는 경우가 있다. 내용의 응집성이 없는 것은 아니기 때문에 EDU로 태깅은 하지만 다른 레이블로는 태깅할 수 없어 별도로 기타 관계(unstated-relation)라는 요소를 마련하였다.

표 2. NC-RIKS (2016) 기본 구조

관계	특징	구분	유형 개수 (총 30가지)
제시관계	텍스트의 핵심 정보를 긍정적으로 고려하거나, 믿거나 받아들이고자 하는 성향을 증가시킴.	[핵(N)] 핵심 정보 [위성(S)] 제시적 기능	10가지
주제관계	연결된 내용 간의 관계적 정보를 인식시킴.	[핵(N)] 중심 내용 [위성(S)] 부차적 내용	15가지
다중핵관계	두 개 이상의 단위가 내용상 대등하게 연결되어 있음.	[핵(N+N+...)] 모든 연결 단위	4가지(연결, 대조, 연속, 다핵재진술)
기타관계	다른 관계에 포함되지 않는 경우	[핵(N)] 주요 내용 [위성(S)] 부차적 내용	1가지 (기타관계)

분류 체계를 정제하는 작업과 위계 구조와 레이블(label) 태깅 작업은 함께 시행되었다. 또 분류 체계 초안을 토대로 태깅 및 검수 과정에서 수정을 거쳤다.

4. 태깅 방법

4.1. 태깅 제약 조건

선행 연구는 문장 내에 EDU가 여러 개 존재할 수 있다고 상정하나, 한국어는 조사와 접속사 등의 쓰임이 명확하게 구분되지 않으므로 이에 대한 지침이 필요하다.

본 연구에서는 종결 어미와 마침표, 느낌표, 물음표 등의 문장 부호를 기준으로 문장을 분할하였다. 이런 과정은 자동으로 처리가 가능하다. 그리고 문장이 여러 개 절로 구성된 경우나 인용절을 포함한 복합문, 또 명사절, 관형사절, 부사절을 포함한 복합문의 경우에 문장 내의 절을 따로 분할하지 않는다. 한편, 제목과 소제목, 원문 하이퍼링크, 작성자, 작성날짜 등은 문장이나 절이 아니더라도 하나의 담화 단위로 취급하여 분할하였다.

수사 관계 연결 시에는 제약을 두어 기준을 마련하였다. 문장 내 요소는 문장 내 요소끼리만 연결된다는 문장 섬 제약과 인접한 단위끼리만 연결이 가능한 인접성 제약이 그것이다. 이러한 제약 조건을 통해 EDU 분할 시 문장 내의 요소는 문장 밖의 요소와 핵-위성, 다중핵 관계를 맺을 수 없다.

4.2. 코퍼스 선정

코퍼스 구축을 위하여 작업할 데이터를 선정하였다. 텍스트에 주석을 직접 달아야하므로 세부 주제와 문장 길이 등을 고려하여 총 760편을 선별하였다. 문서 선정

시에는 균형을 위하여 문서 길이와 주제를 모두 고려하였다. 전체 코퍼스 구성과 규모는 [표 3]과 같다.

표 3. 코퍼스 구성 및 규모

주제	세부	문서 길이 (문장 수)				계
		5-10	11-20	21-30	31-45	
스포츠	야구	120	120	120	120	480
경제	경제 일반	10	10	10	10	40
	국제 경제	10	10	10	10	40
	금융 증권	10	10	10	10	40
	부동산	10	10	10	10	40
	산업/ 기업	10	10	10	10	40
	생활 경제	10	10	10	10	40
	취업/ 창업	10	10	10	10	40
	누계	70	70	70	70	280

4.3. 태깅 절차

태깅 수행에 앞서 대상 텍스트를 숙지하고 텍스트의 주요 내용을 파악하였다. 그 후 텍스트 내에서 관계를 형성하는 각 부분의 범위를 표시하고 관계성에 따라 텍스트 내 각 부분을 연결하였다. 관계의 종류를 표지한 연결 관계를 rstWeb 툴[2]을 이용하여 작업자 3명이 동일한 절차에 따라 원문을 분석하였다. 작업자는 rstWeb 서버에 접속하여 작업하고 논의를 통해 합의된 결과로 태깅하였다. 그리고 완성된 결과물에 대해서는 면밀한 검수 및 수정을 거쳤다.

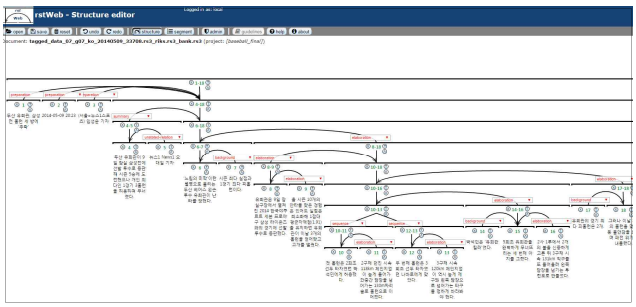

그림 2. rstWeb 실행 화면

작업 시 가장 큰 우려사항은 연구자의 주관에 의해 판단될 소지가 있으며 완벽하게 통일된 분석이 어렵다는 점이다. 따라서 작업자 간 여러 번의 협의와 피드백을 거쳐 가이드라인을 작성하였고 가이드라인 역시 검수과정을 거쳤다.

이와 같은 과정을 거쳐 분석과 학습이 용이하도록 마크업 언어(Markup Language)로 구조화된 코퍼스를 구축하였다. RST 태깅을 위해 도입한 웹 기반 rstWeb 툴은 작업과 검증의 효율성을 높이고, 태깅 오류를 최소화하는 데에 크게 기여한 것으로 평가하였다.

태깅 작업은 다음과 같은 단계별 절차로 수행되었다.

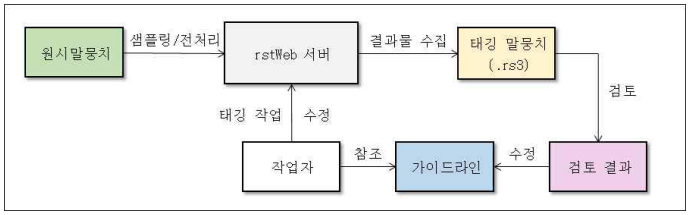

그림 3. 태깅 작업 절차

5. 분석 결과

5.1 각 성분 분포 비율

사용 빈도에 따른 분포도를 [그림 5,6]과 같이 나타낼 수 있다. 제시 관계와 주제 관계에서는 각각 도입과 정교화 관계가 가장 높은 빈도를 보였으며, 기사문의 장르적 특성상 특정 레이블에 치중된 경향을 보였다.

5.2 중심 문장 구조

RST tree에서 중심문장은 시각적으로 쉽게 판별할 수 있다. 최상위 구조로서 뿌리(root) 부분에 해당하고, 내부적으로는 전후 제시 관계와 주제 관계로부터 동시에 화살표를 받는다. 다만 중심 문장 후보가 여러 개인 경우에는 보다 대표성을 띤 문장을 중심 문장으로 판단하며, 문장 간의 관계가 대등한 경우 먼저 나온 문장이 중심문장일 가능성이 높다. 대표성을 비교하는 기준은 해당 문장이 포함하고 있는 요소들이 본문 전체의 내용을 얼마나 잘 반영하고 있는지에 달려있다.

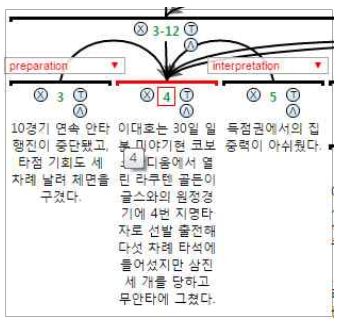

그림 4. 중심 문장 구조의 예

중심문장 후보 문장들이 인접해 있다면 중심문장으로 선택한 한 문장을 핵으로 하고 나머지 문장을 위성으로

2) https://corpling.uis.georgetown.edu/rstweb/info/

하여 서로를 재진술하는 관계로 나타난다. 후보 문장이 인접해 있지 않고 떨어져 있는 경우 중심문장을 기준으로 전체 구조를 연결한 뒤 남은 후보 문장은 요약 관계로 연결된다.

텍스트의 성격과 목적에 따라 저자의 의도나 문장의 내용이 바뀌기 때문에 레이블 간 조합이 일관되게 나타나는 경우는 적었다. 구축한 코퍼스에서는 [그림 5,6]과 같은 전체 레이블별 빈도가 관찰된다. 신문 기사의 구조적 특성상, 도입으로 제시한 뒤 정교화 관계로 구체화시키는 연결이 가장 많이 등장하는 것으로 해석된다.

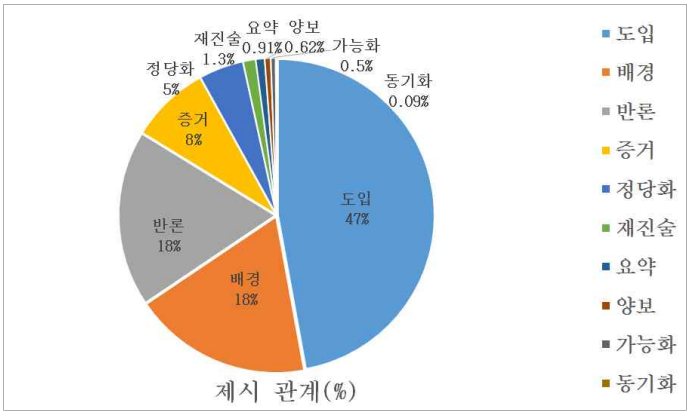

그림 5. 제시 관계 레이블 분포

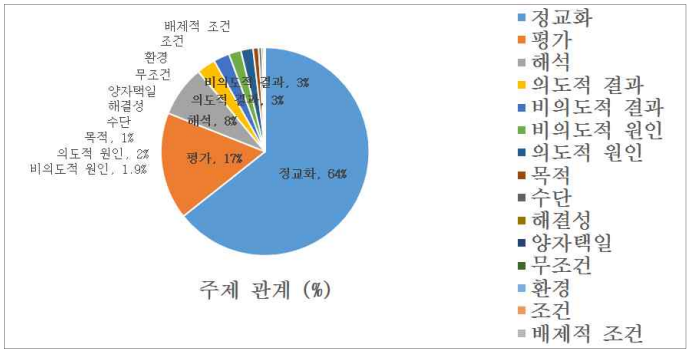

그림 6. 주제 관계 레이블 분포

5.3 위계적으로 상위에 있는 문장의 특징

이러한 문장은 RST tree 구조적으로 상대적으로 위에 위치하며 가능한 의존 관계를 포함하고 위아래 많은 화살표로 연결되어 있다. 상위에 있는 문장 중에서 대체로 제시 관계가 선행하고, 주제 관계로 설명을 덧붙인 문장은 의미적으로 중심 문장인 경우가 많다. 이와 대조적으로 전체 문장 연결 수가 적고 고립될수록 중심 내용과 관련성이 적은 문장이다.

RST tree에서 최상위 중심 문장은 그 자신은 아무런 태깅이 붙지 않는다. 이들 문장은 레이블이 없고, 구조상 다른 EDU와 가장 많이 연결되는 특징을 보인다.

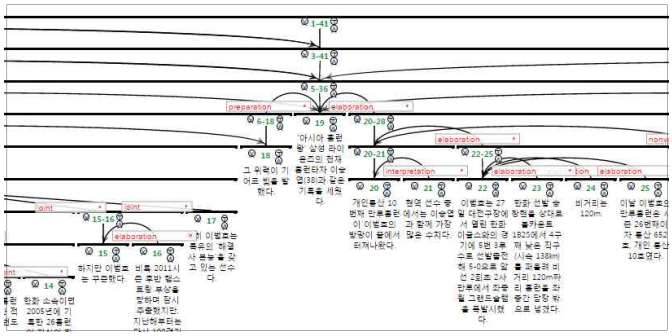

그림 7. 상위 중심문장 구조의 예

5.4 신문기사의 장르적 특성

신문 기사의 텍스트 구조는 크게 표제와 부제, 서두(lead)-본문(body)-결말(ending)의 형식을 띤다. 서두에서는 주로 핵심 정보를 요약하여 주제를 제시하거나 개괄하는 내용이 등장하고, 본문에서는 부가 정보를 제시함으로써 상세하게 설명하는 역할을 한다. 그리고 결말 부분은 주로 서두 내용에 대한 확장 및 심화를 위해 추가적인 정보를 제시한다.

하지만 일부 기사에서는 보도 기사의 마무리가 본문의 주제를 벗어나지는 않되, 글의 중심 내용과는 큰 관련성이 없거나 부가적인 정보로 분위기가 전환되는 경우가 있다. 이런 경우를 위해 설정한 것이 기타 관계로서, 본문 전체에 대한 위성 구조로 연결되는 형식을 띤다.

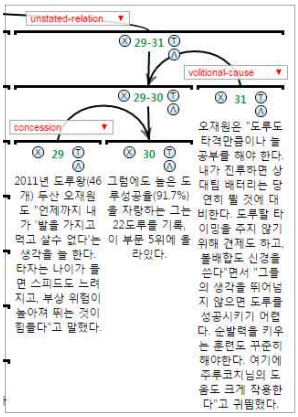

그림 8. 기타관계 예시

한편, 본문 내용이 길어 여러 단락으로 구성될 때 단락의 연계를 위해 접속어 대신 핵심 키워드나 문장을 일종의 표제어로 제시하거나, 복수의 단락끼리 묶고 제각기 부표제를 붙이는 경우도 관찰된다. 전체 구조에 따라 부표제를 글 내부의 또 다른 주제의 도입 구조로 연결하여 상위의 구조와 연결할 수 있으며, 각기 대등한 중심 키워드인 경우 다중핵 관계로 연결된다.

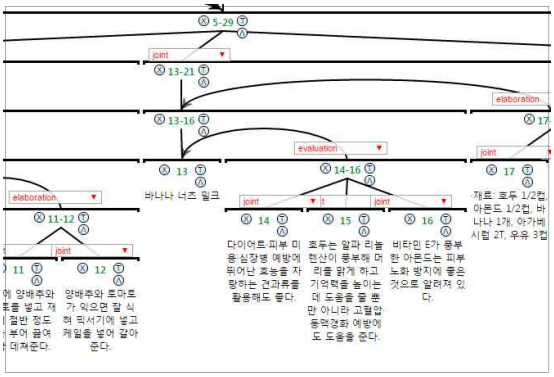

그림 9. 키워드 표제 예시

6. 결론 및 향후 연구

본 연구는 760개 기사문에서 구축된 코퍼스의 1만 5천여 개 이상의 EDU에서 다양한 형식의 화행과 이를 수반하는 다양한 담화상의 요소와 전략을 분석하는 것을 목적으로 하였다. 수사구조이론 외에 화행이론, 담화분석이론 등의 화용론의 주요 이론을 기초로 각 EDU로 나타나는 화행을 정의하고, 한 유형의 발화 형태가 실제 언어의 사용에서 다양한 유형의 화행으로 구현되는 과정과 원리를 살펴보고자 하였다. 또 실제 기사문 텍스트에서 각 수사 구조가 사용되는 분포와 빈도를 연구함으로써 자료 기반 귀납적 연구와 이론 연구를 보완한 한국어 텍스트에 적합한 언어 수행 연구를 시도하고자 하였다.

이번에 구축한 코퍼스는 공시적(synchronic)인 균형 코퍼스이자 학습 데이터로 사용가능한 주석 코퍼스로서 폭넓은 활용 가능성을 갖고 있다. 구축된 코퍼스와 연구 결과를 활용하여 향후 한국어의 다양한 수사 구조에서 실제 사용되는 양상을 분석하여 한국어 텍스트의 장르별 특성을 밝히고 각 빈도와 분포를 고려한 응집성이 높은 문장 구조 추천 혹은 작문 기법 예시 등을 보여줄 수 있을 것으로 기대한다. 뿐만 아니라 한국어 텍스트 자동 요약 시스템과 한국어 문장 자동 배열 시스템 개발을 위한 학습 데이터로도 활용이 가능할 것이다. 이밖에도 관계 유형에 따라 독자의 행동을 유도하는 문장을 학습시키는 데에도 보다 유용할 것으로 판단된다. 따라서 향후 국내 야구 및 경제 뉴스 기사 주석 코퍼스(tagged corpus) 뿐만 아니라 다른 주제와 장르의 텍스트에 대해서도 구축할 필요가 있을 것으로 보인다.

참고문헌

[1] Mann & Thompson, “Rhetorical Structure Theory: Toward a functional theory of text organization”, Text, 8(3), 243-281, 1988.

[2] Jurafsky & Martin, Speech and Language Processing, 2nd edition, Pearson Education Ltd., 2009.

[3] 남기택, 최승기, “수사구조이론을 활용한 현대시 텍스트 분석-「진달래꽃」과 「님의 침묵」을 중심으로”, 한어문교육, 제28집, pp.155-179, 2013.

[4] 남길임, “이론으로서의 말뭉치 언어학에 대한 연구 현황과 쟁점”, 한국어의미학, 제46권. pp.163-187, 2014.

[5] 송경화, 강범모, “신문 기사의 언어 사용 양상”, 인지과학, 제17권, 제4호, pp.255-269, 2006.

[6] 황희선, “RST를 활용한 연결어미 사용 문맥 연구-칼럼에서의 대등 연결어미 ‘-고, -지만, -(으)며’를 대상으로”, 어문논집, 제63권, pp.169-190, 2015.

[7] 김성옥, “수사구조이론에 기초한 영한번역 연구”, 언어학연구, 제17권, 제3호, pp.1-24, 2012.

[8] 이선영, “수사구조이론을 활용한 논증 텍스트 분석 방안”, 한국작문학회, 제25권, pp.101-126, 2015.

[9] 이소현, “수사구조 이론에 기반한 “-지만”의 의미 연구“, 언어와 언어학, 제66권, pp.301-323, 2015.

[10] 윤석민, “RST와 국어의 텍스트 분석”, 텍스트언어학, 제1권, pp.127-167, 1994.

[11] 서성교, “논증 텍스트와 비논증 텍스트의 수사구조”, 언어학, 제11권, 제4호, pp.39-58, 2003.

[12] 이해윤, 전수은, “텍스트 유형별 구조 비교분석-수사구조이론을 기반으로”, 텍스트언어학, 제23권, pp.231-254.

[13] 김재봉, “수사구조이론을 활용한 요약 전략과 적용”, 한국언어문학, 제37권, pp.25-40, 1996.

[14] 이원표, “신문 사설에서의 직접 인용: Bakhtin의 ‘대화성(dialogicality)’ 관점에서의 분석”, 담화와 인지, 제12권, 제2호, pp.117-151, 2005.

[15] 정여훈, “수사구조이론과 한국어 텍스트 분석의 실제”, 언어사실과 관점, 제32권, pp.261-288, 2013.

[16] van der Vliet, N., “Syntax-based discourse segmentation of Dutch text”, In *15th Student Session, ESSLLI*, pp. 203-210, 2010.

[17] Iruskieta, M. & Zapirain, B., “EusEduSeg: A dependency-based EDU segmentation for Basque”, *Procesamiento del Lenguaje Natural*, 55, pp.41-58, 2015.

[18] Iruskieta, M., Dias de Ilarraza, A. & Lersundi, M., “Establishing criteria for RST-based discourse segmentation and annotation for texts in Basque”, *Corpus Linguistics and Linguistic Theory*, 11(2), pp.303-334, 2015.