

# 텍스트 난이도에 따른 핵심 어휘와 그 관계망 변화에 대한 연구\*

이삼형(한양대)\*\* · 길호현(서원대)\*\*\*

---

## <차 례>

---

1. 머리말
  2. 앞선 연구의 검토
  3. 연구의 대상 및 연구 방법
  4. 난이도에 따른 핵심 어휘와 그 관계망의 변화
  5. 맺음말
- 

**【벼리】** 이 글은 글의 난이도에 따라서 글의 핵심 어휘들이 관계를 맺는 양상이 변함을 밝히는 논문이다. 난이도에 따른 분석 대상 텍스트로는 국가수준 학업성취도평가의 초등, 중등, 고등 읽기 지문 및 진공자를 대상으로 하는 글을 선정하였다. 분석은 R 프로그램을 이용하여 텍스트 마이닝을 수행하였으며, 단어의 출현 빈도와 상관관계를 산출하고 단어 사이의 네트워크 구조를 시각화했다. 분석 결과 텍스트의 난도가 높아질수록 핵심어의 군집성이 약화되었으며 핵심어들 사이의 연결 관계가 다층적으로 형성되고 네트워크의 범위가 넓어졌다. 이 연구는 텍스트 마이닝을 활용하여 텍스트의 구조를 시각적으로 드러냈다는 점에서 의의가 있다.

\* 주제어 : 텍스트 난이도, 주제어, 핵심 어휘, 텍스트 마이닝, R, 사회 네트워크 분석

---

\* 이 논문은 2015년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2015S1A5A2A01014617)

\*\* 이삼형(제1저자), \*\*\*길호현(교신저자)

## 1. 머리말

우리가 만나는 담화 상황에서 어떤 대상, 사건, 작품 등을 전달하거나 설명할 때, 텍스트의 난이도가 문제가 되는 경우가 많다. 예를 들면 다음과 같은 담화 상황이 그것이다.

어린이 신문이나 어린이 잡지의 기사

동화 <홍길동전>을 초등학생에게 읽히는 경우

학습 부진아에게 학습할 내용을 쉽게 설명해야 하는 경우

위의 예에서 알 수 있는 바와 같이 텍스트의 난이도 조정은 일상적인 담화 상황은 물론이고, 사회 문화적 현상 및 교육적 상황에서 일어나는 매우 일반적인 문제이다. 그러나 어떠한 텍스트가 쉽거나 어렵다고 학문적으로 판단하는 것은 용이한 문제가 아니다. 쉬운 텍스트와 어려운 텍스트의 차이는 무엇이며 그것을 판단하는 기준은 무엇인가?

전통적으로 국어교육학 특히 독서 교육학에서는 이러한 문제를 텍스트 난이도(difficulty) 또는 이독성(readability)이라는 주제로 연구하여 왔다. 텍스트의 난이도는 텍스트의 어려운 정도를 나타내는 상대적 개념이며, 이독성은 텍스트의 읽기 쉬운 정도를 나타내는 개념으로 그 연구 주제 및 목적은 유사한 편이다.

지금까지 이독성에 대한 연구는 크게 두 가지 측면에서 접근하여 왔다. 하나는 독자적 요인이다. 독자의 인지적 능력과 지적 수준에 적합한 텍스트는 무엇일까에 대한 관심이 그것이다. 다른 하나는 텍스트 내적인 특성 즉, 어휘, 문장, 구조 등과 이독성과의 관련에 대한 연구가 있는데 지금까지 이독성 공식은 주로 후자의 측면에서 구안되었다고 할 수 있다.<sup>1)</sup>

1) 기업에서도 이독성을 위한 프로그램이 개발되어 있기도 하다. 우리나라의 경우, ‘(주)날말’의 어휘정보처리연구소에서는 개인의 독서능력과 도서의 이독성 평가를 위해 ‘우리말 독서력 평가(READ-LQ Assessment)’를 개발하였고, 어휘의 의미와 의미 간의 관계

텍스트 이독성이 어휘의 난이도에 영향을 많이 받는 것은 당연한 일이다. 기본 어휘나 1차 어휘 등 어휘는 그 의미역에 따라 다양한 난이도를 지니게 된다. 그러나 어휘의 유무 외에 특정 텍스트 내의 어휘가 다른 어휘와 어떠한 사회 관계망을 형성하는지에 대한 반영은 미흡하다.

한편 이독성은 문장의 길이와 통사적 복잡성에 의해서 결정되기도 한다. 나아가 텍스트적인 측면도 이독성에 많은 영향을 미친다고 보고 있다. 이는 최근 텍스트에 대한 연구가 활발하게 이루어진 것에 힘입은 바가 크다. 주로 텍스트의 구조가 이독성과 맺는 관련성과 텍스트성이 이독성에 미치는 영향에 대해서 관심을 갖기 시작하였다. 그러나 어휘만큼 명확히 판별하기가 어렵고 그 차이를 프로그램을 통해 기술하기는 현재로서 쉽지 않은 상황이다.

텍스트 난이도 또는 이독성에 대한 연구는 특정 텍스트의 난이도(이독성)를 판별하는 공식을 개발하는 것이 주류를 이루고 있다. 이에 대한 연구는 Gray and Leary에 의해 포괄적으로 진행되었는데 이들의 공식은 768개의 Dale의 어휘 목록을 기준으로 판별한 ‘어려운 단어의 수’, ‘이해 점수의 평균’, ‘전치사 구의 수’ 등의 요인을 고려하였다(최천택, 1995:323). Gray and Leary의 공식 이외에도 Lorge 공식, Flesch의 RE 공식, Dale-Chall 공식, Fog Index 등의 공식이 알려져 있다.

현재까지 제안된 이독성 공식만도 수백 개가 넘는다고 알려져 있으며 지금도 이독성을 측정하는 기준이나 방법들이 새롭게 제안되고 있는 실정이다. 물론 이러한 이유는 지금까지 제안된 이독성 공식이 만족스러운 수준에 이르지 못하였기 때문이다. 그만큼 이독성에 관한 연구가 아직은 미진한 수준에 있다.

여기서 우리가 주목할 바는 텍스트 난이도의 교육적 의의이다. 대표적으로 장미경(2012)은 텍스트 난이도의 연구가 학습자들의 텍스트 이해 정도를 판

---

등 어휘와 관련된 지식수준을 평가하기 위해 ‘우리말 어휘력 평가(Korean Vocabulary Assessment)’를 개발하여 서비스를 제공하고 있다. 미국에서도 일찍이 이와 유사한 서비스를 제공하는 회사들이 있으나 시장 및 현장 교육의 반응은 미미한 편이다.

단함으로써 학습자들의 읽기 능력을 고려한 적절한 텍스트 선정과 개발, 재구성에 크게 도움이 된다고 보았다. 또한, 텍스트 난이도에 대한 연구가 읽기 텍스트의 난이도를 다각적이고 종합적인 관점에서 분석하려는 유의미한 시도라고 평가하였다. 서혁(2011)은 텍스트 난이도의 판단 기준을 정밀화하는 연구가 앞으로 학습자의 수준에 맞춘 체계적인 읽기 교육을 수행하는 데 도움이 될 것으로 보기도 하였다.

텍스트 난이도에 대한 연구는 국어교육에서만 유용한 것이 아니라 외국인을 대상으로 하는 한국어교육 분야에서도 매우 시급한 연구이다. 영어교육 분야에서는 이 분야에 일찍이 관심을 기울여 난이도 분석을 위한 다양한 소프트웨어와 난이도 공식들이 개발되었으며, 이를 통해 학습자에게 적절한 텍스트 선정 및 교수를 위한 난이도 연구가 활발히 진행되고 있다. 한국어교육학 분야의 경우, 학습자가 생산한 한국어 텍스트의 난이도 분석 방안의 마련은 시급한 문제이며, 교사들이 쉽게 활용할 수 있는 소프트웨어 또한 매우 부족한 실정이다. 지금 불고 있는 한국어 학습의 열기를 지속시키기 위해서도 학습자들에게 적절한 텍스트를 제공하여 효과적인 한국어교육이 이루어질 수 있도록 이 방면에 연구가 요청된다.

본 연구의 주제와 관련이 있는 텍스트 난이도는 외국에서는 이미 오래 전인 1930년대부터 연구되기 시작하였다. 우리나라에서도 영어교육 및 국어교육 분야에서 많은 관심을 보인 분야이다. 이런 점에서 본 연구의 주제 분야는 새로운 주제라고 보기 어렵다. 오히려 오랫동안 연구되어 왔으나 아직까지 완벽하게 해결되지 않은 오래된 연구 주제라고 할 수 있다.

본 연구는 연구 주제 그 자체보다 연구 주제에 접근하는 방법론에서 차별점을 추구하고자 한다. 지금까지의 연구들은 어휘를 중심으로 하거나, 통사적인 요소, 텍스트의 구조 등을 활용하여 텍스트의 난이도의 정도나 이독성을 측정하는 공식을 만드는 데 주안점을 두었다. 본 연구가 이전의 연구들과 차별되는 점은 다음과 같다.

먼저 본 연구는 실제로 이루어지는 담화 현상을 기반으로 텍스트의 난이도

의 차이가 반영된 텍스트의 특성에 관심을 갖는다. 즉 본 연구는 이독성 공식을 만들고 특정 텍스트가 어느 수준에 해당하는지를 판별해 내는 것에 관심을 갖는 것이 아니다. 지금까지의 이 방면에 대한 연구는 실용적인 목적을 너무 앞세운 경향이 있다. 본 연구는 이독성 공식과 무관하게 이독성과 관련된 현상이 현재 많이 이루어지고 있다는 사실에 주목한다. 아울러 그러한 현상은 전문가적인 안목을 바탕으로 이루어진다고 가정한다. 위에서 예를 든 이독성 관련 현상들에서 텍스트를 생산하는 주체들은 출판사 편집자, 교과서 편찬자, 신문 기자와 편집자, 문학 연구자 등 나름대로 전문가 집단이다. 본 연구는 전문가 집단에 의해서 만들어진 텍스트들을 분석 비교하여 텍스트 난이도에 따라 나타나는 특성을 찾아내고자 한다. 물론 이러한 연구가 많이 이루어지면 새로운 방법으로 이독성을 판별할 수 있는 단계까지 나아갈 수 있을 것이다.

이를 위해 본 연구는 디지털 기술의 발전에 의해 제안되고 있는 빅 데이터 분석 방법의 도움을 받으려고 한다. 지금까지 이독성 공식은 언어학에 바탕을 두고 접근하였다. 어휘의 난이도는 어휘론, 문장의 길이와 복잡성은 통사론 등에 기반하고 있는 것이 그것이다. 연구 방법이나 도구들이 발전하면 기존의 방법을 통해서도 알 수 없는 새로운 사실들을 발견할 수 있는 것은 주지의 사실이다. 지금까지 이독성 공식이 만족스럽지 못하다는 점을 생각하면 이 방면에도 새로운 방법의 연구 방법이 동원되어서 새롭게 연구되어야 할 필요성은 충분할 것이다.

본 연구에서 활용하고자 하는 텍스트 마이닝은 빅데이터 처리 방법의 하나로 비정형 데이터인 자연언어를 분석하고 처리하기 위해서 여러 방면에 활용되고 있다. 예를 들면 회사에서 이용하는 경우, 자기 회사의 상품에 대해 언급한 내용들을 트위터, 블로그 등 인터넷 공간에서 대량으로 수집하여 시간적으로 어떻게 변화해 갔는지를 분석하고 그 정보들을 상품의 생산 및 회사의 경영에 반영한다. 본 연구도 이와 유사한 구조를 갖는다. 즉, 독자를 고려하여 만들어진, 난이도가 다른 여러 수준의 텍스트들을 수집하여 텍스트 마

이닝 기법을 활용하여 핵심 개념들을 추출하고, 핵심 개념들과 관련을 맺는 어휘들이 무엇이며 그것들이 어떤 군집을 형성하고 있는지 등을 분석한다. 이를 통해 텍스트의 난이도에 따라 그것들이 어떻게 변화하고 있는지를 파악할 수 있다. 이러한 분석은 위에서 예를 든 회사들이 빅 데이터를 분석해서 유용한 정보를 획득하듯이 난이도에 따른 텍스트의 변화 양상들에 대해 유용한 결과들을 얻을 수 있을 것으로 기대한다.

정리하자면, 본 연구는 텍스트의 난이도에 영향을 미치는 요인들 중 어휘나 문장 구조의 차원을 벗어난 다른 요인들이 있을 것이라는 전제 아래, 자연어로 구성되어 있는 텍스트의 어휘 및 각 어휘들이 맺고 있는 관계망을 텍스트 마이닝 기법을 사용하여 분석함으로써 텍스트의 난이도에 새롭게 접근할 수 있는 가능성을 타진해 보는 것을 목적으로 한다.

## 2. 앞선 연구의 검토

본 연구와 관련된 선행 연구는 두 가지 방향에서 살펴볼 수 있다. 먼저 텍스트의 이독성과 관련된 연구를 통해 텍스트의 난이도에 영향을 미치는 요인들의 연구 현황을 밝힌다. 또한 국어교육 내에서 텍스트 마이닝을 활용한 연구들을 살펴 연구 방법의 시사점을 도출한다.

### 2.1. 텍스트 난이도 관련 연구

텍스트의 난이도와 관련하여 영어로 된 텍스트는 Flesch나 Dale-Chall, Flesch-Kincaid 등의 지수로 표시되는 것이 일반적이다. 그런데 이러한 지수는 기본적으로 어휘의 등급에 기반을 둔 것으로 한국어 텍스트에 곧바로 적용하는 것은 불가능하다. 이에 따라 국내에서도 텍스트의 난이도, 복잡도 또는 이독성을 나타내려는 연구가 진행되었다. 정옥년(2003)은 어휘 외에도 글씨 크기와 배치, 예측 가능성, 삽화, 개념의 복잡성 등의 요소를 고려해야 한

다고 하면서 Rog & Burton의 10단계 텍스트 수준을 제시하였다. 다만 이는 학령별 텍스트 수준에 대한 기준으로 활용할 수 있지만 정성적 평가에 의존할 수밖에 없다는 한계가 있다. 윤창욱(2006)은 어휘의 등급을 다양하게 제시하는 것이 아니라 기초 등급에 해당하는 어휘의 목록을 제시하였다. 이 어휘 목록은 이후 이독성 연구에 다양하게 활용되면서 이독성 연구의 발전에 큰 기여를 하였다. 다만 어휘 목록에서 동음이의어가 구분이 안 되어 있는 등 지속적으로 어휘 목록을 보완할 필요가 있다. 이나래(2010)는 대학수학능력 시험 언어영역의 지문에 사용된 문장을 분석하여 지문 텍스트의 통사적 복잡도를 수치로 나타내었다. 이는 텍스트 난이도의 분석 대상을 어휘에서 문장으로 확대했다는 의의가 있으나 모든 문장의 구문구조를 직접 분석해야 한다는 점에서 활용성의 한계가 존재한다. 서혁(2011)과 서혁 외(2013)는 텍스트 복잡도라는 용어를 제시하면서 텍스트 복잡도를 판단하기 위한 등급별 어휘 목록을 자체적으로 선정하고 문장의 복잡도를 산정하는 지수를 개발하여 초등학교급과 그 이상 학령에 적용할 수 있는 구체적인 텍스트 복잡도 공식을 제시하였다. 이는 텍스트 난이도에 관련된 최근의 연구 가운데 가장 높은 수준의 완성도를 보여주지만 모든 문장의 복잡도를 직접 계산해야 하는 한계는 여전히 안고 있다. 텍스트의 난이도는 한국어교육 분야에서 그 중요성이 더욱 크다고 할 수 있는데, 장미경(2012)은 한국어 텍스트의 난이도 평가 방안을 매우 구체적이고 체계적으로 제시하였다. 최숙기(2012)에서는 미국 CCSS에서 제안한 텍스트 복잡도 모형을 활용하여, 텍스트의 복잡도를 평가하는 양적 요인과 질적 요인을 제시하였다. 양적 요인은 텍스트 이독성 공식을 통해 산출하고 질적 수준은 텍스트 평가 루브릭을 통해 산정할 것을 제안하였다. 이 연구에서는 텍스트의 질적인 요인을 평가하는 루브릭을 제시했다는 점에서 큰 의의가 있지만 정성적인 요인의 평가는 복수의 전문가에 의해 수행되어야 한다는 점에서 일반인들이 쉽게 활용하는 데에는 한계가 있다.

이상으로 텍스트의 난이도를 평가하기 위한 연구들을 살펴보았다. 텍스트의 난이도를 측정하기 위해 가장 먼저 사용되는 요소인 어휘는 정확한 어휘

등급이 선행되어야 한다. 문장 구조의 요소는 텍스트 난이도에 영향을 미치지만 각 문장의 구조를 전문가가 하나하나 산정해야 한다는 물리적 어려움이 존재한다. 텍스트에 대한 질적인 평가 역시 전문가의 질적인 평정에 의존해야 한다는 한계가 있다. 본 연구에서는 이러한 한계를 극복하고 텍스트 분석을 위한 일정 수준의 자동성과 신뢰성을 확보하기 위한 텍스트 마이닝을 활용하기 위한 가능성을 탐색하고자 한다.

## 2.2. 텍스트 마이닝을 활용한 연구

텍스트 마이닝을 활용한 연구 방법은 문서에 나타난 키워드를 분석하기에 유용하기 때문에 텍스트 자체의 구조를 분석한다기보다는 대량의 문서에서 어떤 경향성을 발견하기 위한 연구가 주로 수행되었다. 노형남(2014)에서는 다수의 정치 연설 텍스트를 대상으로 텍스트 마이닝을 수행하여 어휘의 빈도를 기반으로 하는 워드 클라우드 형태로 산출하고 그를 바탕으로 경향성을 도출하였다. 김선아 외(2015)에서는 미술과와 국어과의 교육과정을 텍스트 마이닝 기법을 활용하여 분석하여 두 교과에의 공통점과 차이점, 그리고 지향점을 분석하였다. 이러한 연구들은 다량의 텍스트를 처리하는 빅데이터 처리 기법의 하나로서 텍스트 마이닝의 방법을 활용한 것으로 정밀한 텍스트 분석과는 거리가 있다. 이슬기·박영민(2017)에서는 중학교 학습자들이 직접 작성한 논설문을 텍스트 마이닝 기법으로 분석하여 품사별 특징을 추출하였다. 이 연구는 기존의 연구와는 달리 다양한 품사의 사용 실태를 제시했다는 점에서 가치가 있는데, 역시 다량의 텍스트를 분석하는 것이 중점이며 어휘 빈도에 기반하고 있다는 한계가 있다. 한편 박치성·이준석(2017)에서는 텍스트 자료 정리 방식 및 절차를 정형화하기 위한 방안을 제시했는데, 텍스트 마이닝 분석의 성과에 전처리가 매우 큰 영향을 미치는 만큼 향후 연구의 완성도 향상을 위해 참고할 필요가 있다.



### 3. 연구의 대상 및 연구 방법

#### 3.1. 연구의 대상

텍스트 난이도에 따라 핵심 어휘 및 관계망에서 어떤 차이가 드러나는지를 분석하기 위해서는 먼저 난이도의 차이가 분명하게 드러나는 텍스트를 분석의 대상으로 선정할 필요가 있다. 그런데 기존에 사용하던 난이도 지수는 앞에서 살펴보본 바와 같이 텍스트에 사용된 어휘나 문장 구조 등의 요소를 통해 산정되는 것이므로 텍스트의 난이도를 객관적으로 나타낸다고 보기 어렵다. 또한 본 연구는 텍스트의 난이도에 따라 달라지는 요소들을 분석하는 것을 목적으로 하고 있으므로 이러한 지수를 텍스트 난이도의 분류 기준으로 설정하는 것은 적절하지 않다.

난이도에 따른 요소의 차이를 비교·분석하기 위해서는 텍스트의 난이도가 보다 객관적인 기준으로 사전에 설정될 필요가 있다. 그렇다면 텍스트의 난이도를 구별하는 기준은 어떻게 설정할 수가 있는가? 본 연구에서는 텍스트의 난이도를 설정하는 주요한 기준으로 예상 독자에 주목하였다. 예상 독자를 고려하면 어휘나 문장 구조는 물론 텍스트의 모든 요소를 독자의 수준에 맞추어 개발하거나 선택해야 하기 때문에 텍스트의 난이도가 예상 독자에 따라 조정되기 때문이다. 그리고 이와 같은 난이도의 조정이 가장 의도적이고 정밀하게 수행되는 장면은 객관적인 평가의 상황이라고 할 수 있다. 특히 전국 단위의 평가에서는 독서 및 각 분야의 전문가들이 참여하여 독자의 수준에 적합한 평가용 지문을 제작한다. 따라서 이러한 지문은 전문가에 의해 난이도의 구별이 이루어졌다고 볼 수 있다.

본 연구에서는 난이도에 따라 구별되는 분석 대상 텍스트로 국가수준 학업성취도 평가의 지문을 선택하였다. 국가수준 학업성취도 평가는 초등학교 6학년, 중학교 3학년, 고등학교 2학년 학생들을 대상으로 각 교과와 하위 영역별 성취 역량을 측정하는 전국단위 평가로, 높은 신뢰도와 타당도, 변별도를

갖춘 공신력 있는 평가이다. 학업성취도 평가 국어에서는 듣기·말하기, 쓰기, 읽기, 문법, 문학 등 국어과의 전 영역을 선택형 및 서답형 문항으로 평가하기 때문에 읽기 능력을 측정하기 위한 지문이 학교급별로 다르게 제시되어 있다. 평가 시행 주체인 한국교육과정 평가원에서 결과 분석 자료로 공개한 문항 정보에서 읽기 지문의 핵심 화제를 정리하면 다음과 같다.

	2009	2010	2011	2012	2013	2014
초 등	지폐 올림픽 지구의 시간 옷 지역축제	쌈소비랑 전래놀이 태양광선 제주고인돌 나트륨섭취	등요 씨앗소유권 체력저하 무인 대여대 나노기술	민화 올바른 비판 문화 방선균 베델 의궤	-	-
중 등	인터넷 실명제 공유자원 헬리콥터 맘 한류 수면부족 온돌	사진사 김수남 성씨와 본관 바비인형 상상축제 강화갯벌	타자기 공병우 가야 화폐 초상화 여성 정치참여	민간숙설의 기원 기업의 사회적 책임	나비박사 석주명 지렁이 지수물가 학급폭력	한센병 이종욱 운석 놀이 손실회피성
고 등	의료영상 물가상승 문화관광 로봇청소기 지도제작	활동적 장년층 전자민주주의 원자력논쟁 백의민족	저작권 스마트패치 로봇 이족보행 독립신문 논조	오케스트라 지휘자 이산화탄소 감축 포지셔닝 전략	정보과잉 무아레 무너 법 규범 해석 추상화	윤리적채식 면역반응 국제개발협력

〈표 1〉 국가수준 학업성취도 평가 읽기 지문 화제 목록

난이도에 따른 핵심 어휘와 관계망의 변화를 분석하기 위해서는 학교급 외에 다른 요소들의 간섭을 최대한 배제할 필요가 있다. 이에 본 연구에서는 초 등-중등-고등이 유사한 화제를 다루고 있는 지문들을 하나의 군으로 묶어서 1차적인 비교의 대상으로 선정하였다. 또한, 글의 갈래가 이러한 차이에 영향을 미치는지 살펴보기 위해 논설문 갈래에 해당하는 지문도 하나의 군을 구성하였다. 다만, 평가용 지문이 대부분 설명문으로 구성되는 관계로 논설문 군의 경우는 화제의 동일성을 유지하지는 못하였다. 아울러, 평가용 지문이

고등학생까지만을 대상으로 한다는 한계를 극복하기 위해, 해당 화제와 유사한 화제로 구성된 글 중 성인 또는 전공자를 대상으로 한 글을 선정하여 함께 비교하였다. 최종적으로 분석을 진행한 텍스트의 화제는 다음과 같다.

	A군 설명문 (미술)	B군 설명문 (경제)	C군 설명문 (사회)	D군 논설문
<b>초등 (01)</b>	민화(2012)	지폐(2009)	지역축제(2009)	올바른 비판 문화(2012)
<b>중등 (02)</b>	초상화(2011)	지수물가(2013)	싱싱축제(2010)	강화갯벌(2010)
<b>고등 (03)</b>	추상화(2013)	물가상승(2009)	문화관광(2009)	전 자 민 주 주의 (2010)
<b>전공 (04)</b>	정물화	소비자물가지수	리우 카니발	통상전쟁

〈표 2〉 분석 대상 텍스트 화제 목록

이 연구는 엄밀한 통계적 방법에 의해서 새로운 사실을 규명한다기보다는 텍스트의 난이도 분석을 위해 텍스트 마이닝을 적용할 수 있는 가능성을 탐색하는 데 의의를 두고 있으므로 테스트의 선정 역시 양적인 측면이나 대표성을 고려하여 엄정하게 이루어진 것은 아니다. 이 부분에 대해서는 후속 연구를 통해 유의미한 결과를 산출하고 이를 검증 및 확산하고자 한다.

### 3.2. 연구 방법

본 연구는 크게 두 가지 방법으로 텍스트 분석을 수행하였다. 먼저 텍스트 마이닝 기법을 활용하여 텍스트의 핵심어를 추출하였으며, 추출된 텍스트의 상관관계를 분석하여 핵심어간의 연결성을 도출하는 사회적 네트워크 기법을 적용하였다. 자세한 연구 방법 및 경과는 다음과 같다.

### 3.2.1. 텍스트 마이닝을 통한 핵심 어휘 추출

텍스트 마이닝은 비정형 데이터인 텍스트를 자연어 처리(natural language processing) 기술에 기반하여 분석함으로써 정보를 추출하고, 가공·분석하는 것을 목적으로 하는 기술이다. 비정형 데이터란 숫자로 확정된 정형 데이터와 달리 텍스트, 인터넷 및 SNS 문서, 음성, 문자 메시지 등 형태와 구조가 다른 데이터를 가리킨다. 텍스트 마이닝은 이러한 텍스트에 사용된 언어를 통계 처리가 가능한 형태로 가공한 후 그 안에서 다양한 정보를 발견한다. 구체적으로 본 연구에서는 앞에서 선정한 지문들을 독립된 txt 파일로 변환한 후 'R' 3.4.3 프로그램을 사용하여 텍스트 마이닝을 수행하였다.

텍스트 마이닝 분석을 성공적으로 수행하기 위해서는 데이터 전(前)처리 작업이 중요한데, 한글 텍스트에서 주요한 전처리 작업은 숫자나 문장 부호 제거, 평가 문항을 위해 사용된 각종 약물이나 밑줄 제거, 문항을 위해 누락되거나 제거된 표현 복원, 오타자 교정 등이 있다. 영어 텍스트의 경우 전처리를 위해서 stopwords 라고 불리는 무의미어(예를 들어 it이나 There 등 문법적으로 자주 나올 수밖에 없는 단어들)를 처리하거나 1개의 철자로 이루어진 단어(예를 들어 관사 'a')를 삭제하는 과정을 수행하게 되는데, 국어는 별도의 무의미어 목록이 없으므로 이는 수행하지 않았다. 다만, 사전 분석 결과 내용과 무관하게 '있다', '하다', '않다'가 빈번하게 등장하였는데 이 단어들은 특정한 의미를 나타내거나 텍스트의 특성을 보여주는 데 기여하지 않는다고 판단되어 빈도 추출에서 제거하였다. 향후 한국어 텍스트에서 불용어로 처리할 어휘 목록에 대해 추가적으로 심도 깊은 논의 및 연구가 필요하다고 판단된다. 국어 텍스트에서는 한 음절로 이루어진 실질 형태소가 다수 존재한다. 예를 들어 초등학생 대상으로 작성된 '민화' 텍스트의 경우 개, 닭 등 민화와 큰 관련이 있기 때문에 단음절 단어를 함부로 삭제해서는 안 된다. 따라서 일반적으로 영어 텍스트 마이닝을 할 때 수행하는 한 글자 단어 삭제 작업은 수행하지 않았다. 또한 영어에서는 전처리 작업의 하나로 활용형 및 변이형을 기본형으로 바꾸어 주는 stemming 작업을 수행하는데 비해 한국어 텍스

트 마이닝에서는 이러한 작업 없이 대부분 명사를 추출하는 방식으로 분석을 진행해 왔다. 이는 한국어 텍스트 마이닝에 주로 사용되는 KoNLP(Korea Natural Language Processing)<sup>2)</sup> 패키지에서 명사를 추출하는 `extractNoun` 함수를 별도로 제공하기 때문인데, 이러한 함수를 이용하면 명사 추출은 용이할 수 있어도 내용 전개에 중요한 역할을 하는 용언들이 누락될 수밖에 없게 된다. 물론 KoNLP에서 제공하는 `SimplePos09`나 `SimplePos22` 함수를 사용하여 형태소 분석을 수행한 후 용언의 어간을 추출하고 각각의 용언들을 모두 기본형으로 바꾸어 분석을 수행할 수도 있다. 그러나 이러한 분석은 전처리 과정에서 점검하거나 수행해야 할 요소들이 증가하게 되어 대량의 문서를 처리하기에는 해결해야 할 문제가 아직 남아 있다. 본 연구에서는 이러한 한계를 극복하기 위해 NLP4kec 패키지<sup>3)</sup>를 사용하였다. NLP4kec 패키지는 입력된 텍스트에 대해서 형태소 분석을 한 후에 용언들을 기본형의 형태로 바꾸어 반환하므로 stemming 작업이 자동으로 수행된다는 장점이 있다.

전처리 작업을 한 후에는 빈도를 기준으로 핵심어 분석을 수행한다. 본래 텍스트 마이닝은 빅데이터 처리 기법의 일환으로 개발된 것으로, 다량의 문서에서 핵심어를 추출하고 문서간의 연관성을 판단하는 데 적합하다. 그러나 이 연구에서는 다량의 문서에서 통계적인 유의미성을 발견하는 것이 아니라 하나의 문서에서 핵심어를 추출하고 단어들 사이의 연관 관계를 도출하는 것이 목표이다. 이를 위해 본 연구에서는 하나의 문서를 문단 단위로 나누어 입력하여 한 문단을 독립된 하나의 문서로 처리하게 함으로써 여러 문서를 처리하는 것과 같은 효과를 발생하게 하였다. 즉, 여러 문단으로 이루어진 한 편의 텍스트를 각 문단별로 독립된 텍스트들의 조합처럼 처리한 것이다.

빈도를 기준으로 핵심어를 추출할 때 가중치의 문제를 해결할 필요가 있

2) Jeon, H.(2013). KoNLP: Korean NLP Package. R Package Version 0.80.1.  
<http://CRAN.R-project.org/package=KoNLP>

3) NLP4kec는 한글 형태소분석기인 ‘은전한닢’을 기반으로 개발되었으며 엑셀로 입력된 다량의 문서를 분석하기에 용이하다. 패키지 및 사용법은 아래 주소에서 받을 수 있다.  
<https://github.com/NamyounKim/NLP4kec>

다. 여러 문서에서 반복적으로 사용되는 표현은 그 용어가 중요한 핵심어이기 때문일 수도 있지만 화제를 드러내기 때문일 수도 있다. 예를 들어, 인터넷 게시판에서 ‘청소기’에 관련된 글을 모아서 분석을 한다고 할 때, ‘청소기’라는 단어는 당연히 가장 빈번하게 등장하는 용어가 되는데 이는 화제를 제시하는 것일 뿐 청소기를 이야기할 때 사람들이 중요하게 생각하는 요소들을 추출하는 데에는 그다지 유용한 정보를 제공하지 못한다. 즉, 여러 문서에서 보편적으로 사용되는 용어는 오히려 그 보편성으로 인해 핵심 정보를 담아내지 못하게 되므로 가중치를 조정할 필요가 있다. 이를 TF-IDF(Term Frequency - Inverse Document Frequency)라고 하는데 여러 문서에서 일반적으로 등장하는 용어의 가중치를 낮추어 실질적으로 중요한 역할을 하는 핵심어들을 추출할 때 사용된다. 그런데 이는 다량의 문서에서 유의미한 핵심어를 발견할 때에는 유용하지만, 본 연구에서처럼 하나의 문서에서 핵심어를 분석하는 경우에는 오히려 중심 화제어가 삭제되는 결과가 도출되기도 하였다. 예를 들어, 초등학생을 대상으로 하는 ‘민화(a-01)’ 텍스트에서 TF-IDF를 반영하여 핵심어를 추출하면 가장 빈번하고 또 중요하게 사용되는 ‘민화’가 사라지게 되어 오히려 핵심어 추출에 적합하지 못한 결과를 얻게 되었다. 따라서 본 연구에서는 TF-IDF를 반영하지 않은 결과를 바탕으로 분석을 진행하였다. 다만, 이 경우 ‘민화’에 영향을 미치는 주요한 요소를 파악하는 데에는 TF-IDF를 반영하는 것이 오히려 유용할 수도 있다. 단일 텍스트의 분석에서 TF-IDF를 반영한 결과와 그렇지 않은 결과를 면밀하게 비교할 필요가 있으나 이는 추후 별도의 연구를 통해 밝히고자 한다.

### 3.2.2. 핵심어 네트워크 분석을 통한 관계도 산출

사회 네트워크 분석 기법은 사람이나 사물과 같은 개체(node)들 간 관계(edge)를 계량적이고 시각적으로 나타내어 개체 간에 연결이 되는 관계의 특성을 분석하는 기법이다(김용학, 2003). 주로 사회과학 분야에서 연구되었던 사회 네트워크 분석 기법은 상품이나 서비스, 정보교환 등 다양한 분야에서

활용이 가능하다(Haythornthwaite, 1996:323-342). 특히 최근에는 카페나 블로그, SNS 등의 분석을 통해 소비자의 성향을 파악하여 새로운 상품이나 서비스를 개발하는 방향으로 다양하게 활용되고 있다.

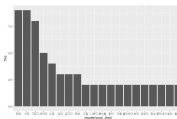
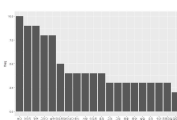
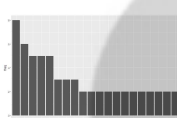
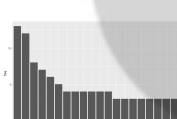
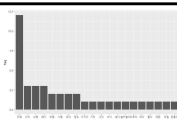
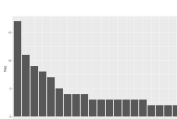
본 연구에서는 하나의 문서가 각 단어들의 네트워크로 이루어진 구조물이라고 가정한다. 그리고 네트워크 분석의 단위는 완결성을 가진 사과의 단위인 문단으로 설정하였다. 이에 따라, 각 문단을 하나의 독립된 문서로 상정하고, 각 문서에 사용된 단어들을 기본형으로 정리하여 추출한 후 각 문단에서 사용된 단어들의 DTM(Document-Term-Matrix)의 행렬을 만들고 이 행렬을 통해 단어들의 상호 상관관계를 산출하는 방식으로 분석을 진행하였다. 또한 각 지문에서 가장 핵심이 되는 단어와 상관관계가 높은 단어들을 별도로 추출하여 관계성을 살펴보았다. 일반적으로 피어슨 상관계수가 0.6 이상이면 높은 상관관계를 인정하므로(성태제, 2014:131) 본 연구에서도 이를 기준으로 하여 상관계수가 0.6 이상인 관계를 시각적으로 표시하도록 하였다.

## 4. 난이도에 따른 핵심 어휘와 그 관계망의 변화

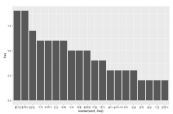
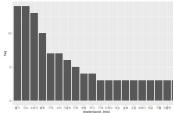
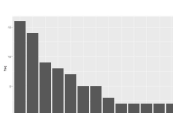
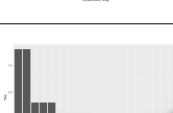
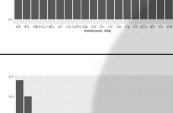
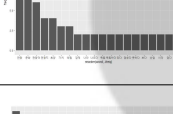
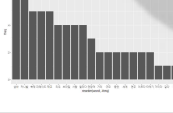
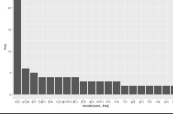
### 4.1. 난이도에 따른 핵심 어휘의 변화

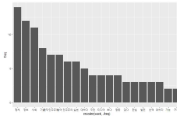
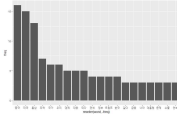
난이도에 따라 핵심 어휘 측면에서 어떠한 변화가 발생하는지 파악하기 위해 텍스트에 사용된 단어의 어휘 빈도 그래프와 기술 통계량을 산출하여 <표 3>에 제시하였다. 그런데 핵심어를 추출하기 위해 단어의 빈도수를 일률적으로 비교하면 당연히 분량이 긴 글에서 더 높은 빈도가 발생할 확률이 높으므로 전체 단어 수 대비 최빈 단어의 비율을 별도로 산정하여 비교할 필요가 있다. 여기에서 전체 단어의 수는 텍스트에 사용된 전체 단어를 중복을 허용하여 헤아린 수와 텍스트에 사용된 단어의 종류로 나누어 생각할 수 있다. <표 3>에서는 이를 '단어 종류/전체 단어 수'의 형태로 표시하였다. 최빈 단어의 빈도 및 비율을 구할 때는 일률적으로 최빈 단어 하나만을 제시하는 것이

아니라 빈도수가 높은 단어들의 특징이 드러날 수 있도록 추출하였다. 또한 비율은 전체 단어 수에 대한 비율로 산정하였다.

부호	제재 및 출전	어휘 빈도 그래프	기술통계량	최빈 단어의 빈도(비율%)
<b>A-01</b>	민화(2012)		문단:4 단어:94/173 (54%) 글자:520	민화: 9(5.2) 서민: 9(5.2) 그리다: 8(4.6) 바라다: 5(2.9) 소망: 4(2.3)
<b>A-02</b>	초상화(2011)		문단:4 단어:108/261 (41%) 글자:801	보다: 10(3.8) 이미지: 9(3.4) 정면: 9(3.4) 그리다: 8(3.0) 측면: 8(3.0)
<b>A-03</b>	추상화(2013)		문단:3 단어:117/197 (59%) 글자:648	황소: 8(4.0) 추상: 6(3.0) 대상: 5(2.5) 피카소: 5(2.5) 활동: 5(2.5)
<b>A-04</b>	정물화		문단:5 단어:217/441 (49%) 글자:1318	작품: 13(2.9) 정물화: 12(2.7) 고호: 8(1.8) 화관: 7(1.6) 의미: 6(1.3)
<b>B-01</b>	지폐(2009)		문단:8 단어:45/104 (43%) 글자:308	지폐: 12(11.5) 나라: 3(2.8) 수명: 3(2.8) 폐기: 3(2.8)
<b>B-02</b>	지수물가(2013)		문단:5 단어:93/220 (42%) 글자:722	물가: 17(7.7) 지수: 11(5.0) 가격: 9(4.0) 오르다: 8(3.6) 소비자: 7(3.2)



<b>B-03</b>	물가상승(2009)		문단:6 단어:120/262 (45%) 글자:840	물가: 9(3.4) 인플레이션: 9(3.4) 상승: 7(2.7)
<b>B-04</b>	소비자물가지수		문단:4 단어:112/210 (53%) 글자:785	물가: 14(6.6) 지수: 14(6.6) 소비자: 13(6.1) 품목: 10(4.8)
<b>C-01</b>	지역축제(2009)		문단:4 단어:63/169 (37%) 글자:479	지역: 16(9.4) 축제: 14(8.2) 고장: 9(5.3) 우리: 8(4.7) 열다: 7(4.1)
<b>C-02</b>	싱싱축제(2010)		문단:6 단어:135/233 (57%) 글자:718	부족: 9(3.8) 축제: 9(3.8) 사람: 4(1.7) 파퓰러니티: 4(1.7) 행사: 4(1.7)
<b>C-03</b>	문화관광(2009)		문단:4 단어:163/269 (60%) 글자:858	관광: 11(4.1) 문화: 10(3.7) 관광객: 6(2.2) 관광지: 4(1.49) 휴양: 4(1.49)
<b>C-04</b>	리우 카니발		문단:6 단어:66/149 (44%) 글자:478	삼바: 7(4.6) 카니발: 6(4.0) 축제: 5(3.4) 퍼레이드: 5(3.4) 학교: 5(3.4)
<b>D-01</b>	올바른 비판 문 화(2012)		문단:4 단어:95/232 (41%) 글자:675	비판: 22(9.5) 상대방: 6(2.5) 든다: 5(2.2)
<b>D-02</b>	강화갯벌(2010)		문단:6 단어:180/350 (51%) 글자:1043	갯벌: 17(4.9) 강화: 9(2.6) 발전소: 9(2.6) 가치: 8(2.3) 조력: 6(1.7)

<b>D-03</b>	전 자 민 주 주 의 (2010)		문단:5 단어:172/344 (50%) 글자:1161	정치: 17(4.9) 정-B보: 12(3.5) 사회: 11(3.2) 기술: 7(2.0)
<b>D-04</b>	통상전쟁		문단:5 단어:238/423 (56%) 글자:1368	중국: 16(3.7) 미국: 15(3.5) 통상: 13(3.1) 무역: 7(1.7)

〈표 3〉 텍스트별 최빈 단어 목록 및 비율

A, B, C, D 네 개의 군에서 초등 단계인 01부터 일반인을 대상으로 하는 04단계까지 어휘 빈도의 특징을 비교하면 다음과 같은 점을 발견할 수 있다.

첫째, 전체 어휘 대비 신규 어휘의 비율은 유의미한 차이가 발견되지 않았다. 난이도가 높아질수록 다양한 표현이 사용되면서 자연스럽게 신규 어휘의 비율이 높아질 것이라고 예상하였으나, 실제 신규 어휘의 비율은 50%를 기준으로 대부분  $\pm 10\%p$  내에 분포하고 있으며 뚜렷한 경향성이 발견되지는 않는다. 즉 난이도에 따른 차이라기보다는 개별 텍스트의 특징이 더 크게 영향을 미치는 것으로 보인다.

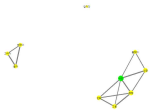
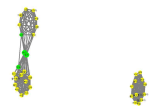
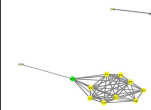
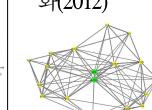
둘째, 난이도가 낮을수록 핵심 어휘가 반복적으로 사용되는 경향이 있다. 이는 최빈 단어의 비율에서 확인할 수 있는데, 각 텍스트에서 가장 자주 사용된 단어들의 빈도를 전체 사용 단어의 수로 나눈 비율 값을 살펴보면 난이도가 낮은 텍스트에서는 핵심적으로 사용되는 단어들의 사용 비율이 높다는 것을 알 수 있다. 이는 난이도가 낮을수록 몇몇의 중요한 표현이나 용어를 반복적으로 사용하고 있으며, 난이도가 높아짐에 따라 중심이 되는 용어들의 사용이 다변화되기 때문으로 해석할 수 있다.

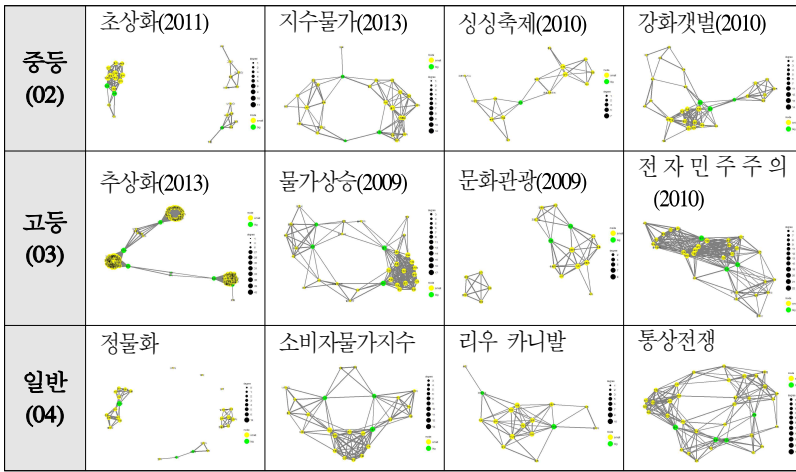
#### 4.2. 난이도에 따른 핵심 어휘 관계망의 변화

앞에서 살펴본 어휘의 빈도를 바탕으로 이번에는 핵심 어휘들이 상호 관계

를 맺고 있는 양상을 시각적으로 구현해 보았다. <표 4>에서는 각 텍스트에 사용된 단어들의 네트워크 구조를 산출하여 한 눈에 비교할 수 있도록 배열하였다. 앞서서도 언급한 바와 같이 네트워크 구조를 산출하기 위해서는 문단별로 사용된 단어들을 일렬로 정렬한 후 이를 행렬로 전환하여 동시 출현 여부와 빈도를 산출한다. 이 과정에서 동시에 출현하지 않는 수많은 단어들은 빈칸으로 존재하는데 이러한 빈칸을 *sparseterm*이라고 한다. 예를 들어, A-01 텍스트에 사용된 신규 단어는 94개인데 이를 행렬로 변환하면 총 8,836(94×94)개의 셀이 생성된다. 이러한 셀을 모두 계산하는 것은 비효율적이기 때문에 빈도가 떨어지는 빈칸을 제거하여 분석의 대상을 한정하게 되는데, 이 연구에서는 95%의 빈칸을 제거하는 것으로 설정하였다. 그런데 A군과 B군의 경우는 사용된 단어의 수가 충분하지 못하여 이를 각각 70%와 90%로 조정하여 적용하였다. 즉 *sparseterm*의 제거 비율을 조정하는 것으로 네트워크 구조에서 분석의 대상이 되는 단어의 수를 설정할 수 있다.

네트워크 구조에서 꼭지점이 되는 단어의 수를 *sparseterm* 비율로 조정한다면, 각 단어들을 연결하는 연결선의 수는 상관계수를 설정하는 방법으로 조절할 수 있다. 단어와 단어 간의 연결 관계에 대한 수와 두께 등은 각 단어 사이의 상관계수에 따라 그려지는데, 본 연구에서는 상관계수 0.6이상인 경우 상관계수에 비례하여 선의 두께가 두꺼워지도록 설정하였다.

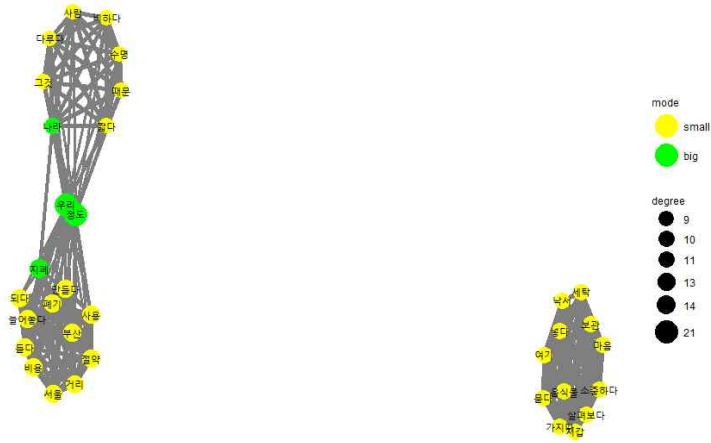
	A군	B군	C군	D군
초등 (01)	민화(2012) 	지폐(2009) 	지역축제(2009) 	올바른 비판 문화(2012) 



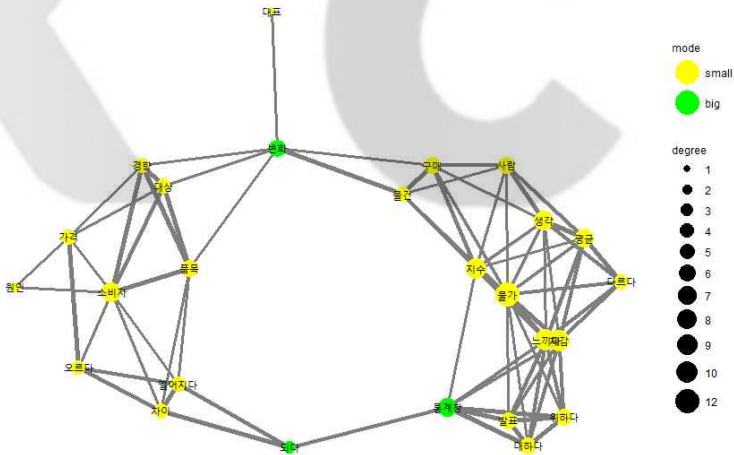
〈표 5〉 분석 대상 텍스트의 사회 네트워크 구조

각각의 그래프에서 중심성이 두드러지는 단어 중 betweenness 지수가 0.9 이상인 단어들은 녹색으로 나타내었으며, 원의 크기는 중심성 지수를 반영하도록 하였다.

예를 들어, 다음의 〈그림 1〉은 B-01 텍스트의 네트워크 구조이다. 이 텍스트는 지폐를 사용하는 방법에 대해서 설명하고 있으며, 어휘 빈도에서는 '지폐'가 독보적으로 높은 빈도를 나타내었다. 그런데 네트워크 구조도에서는 '우리'와 '정도'가 중심을 이루고 있으며, 이 두 단어를 중심으로 각각 두 개의 의미 군으로 나뉘어 상호 연결이 되고 있음을 알 수 있다. 또한 지폐의 사용 방법은 별도의 군으로 독립되어 있는데, 이는 지문 내용 중에서 별도로 수집 구조를 활용하여 지폐 사용 방법을 제시하고 있는 부분에 해당한다. 이 부분은 그 이전의 내용과 의미적으로 공통점이 없는데, 네트워크 구조도에서도 이러한 내용 구분이 시각적으로 잘 드러나고 있음을 확인할 수 있다.



〈그림 1〉 B-01 텍스트의 네트워크 구조

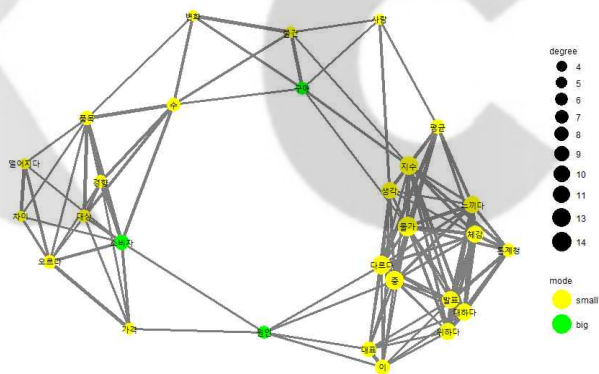


〈그림 2〉 B-02 텍스트의 네트워크 구조

〈그림 2〉는 B-02 텍스트의 네트워크 구조를 나타낸 것이다. 이 텍스트는 초등학교를 대상으로 하는 B-01과는 달리 전체적인 어휘가 분절적으로 나뉘

지 않고 서로 연결되고 있는 것을 볼 수 있다. 네트워크 구조상으로 중요한 단어는 ‘변화’와 ‘통계청’으로 이 두 단어들이 전체적인 내용을 연결해주고 있으며, ‘소비자’와 ‘물가’ 및 ‘지수’가 서로 반대의 영역에서 의미군을 형성하고 있다는 것을 알 수 있다.

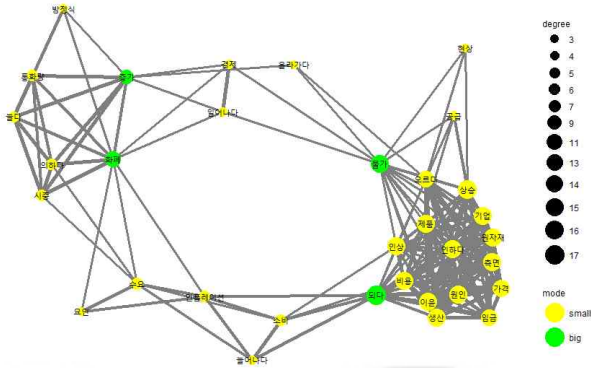
〈그림 3〉은 B-02 텍스트에서 TF-IDF를 반영하여 핵심어를 추출하고 그 결과를 네트워크 구조로 그린 것이다. 그 과정에서 의미적 변별성이 떨어지는 ‘되다’ 역시 불용어로 처리하였다. 〈그림 3〉에서는 ‘소비자’, ‘구매’, ‘원인’ 등이 의미상 구분되는 덩어리들을 연결해 주는 고리 역할을 하는 것으로 나타났다. 〈그림 2〉와 비교하여 〈그림 3〉은 각 단어들이 맺고 있는 의미 관계가 더 긴밀하고 다층적임을 보여준다. 그리고 〈그림 1〉의 대상이 되는 초등 수준의 텍스트와 비교하면 의미 관계의 분절성이 약화되고 연결성이 한층 활성화되었다는 것을 확인할 수 있다.



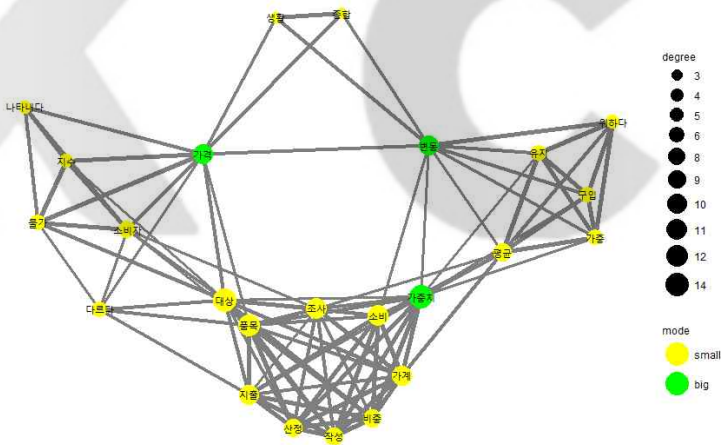
〈그림 3〉 B-02 텍스트에서 TF-IDF를 반영한 네트워크 구조

〈그림 4〉는 B-03 텍스트의 네트워크 구조를 나타낸 것이다. 이 텍스트는 고등학생을 대상으로 작성된 지문인데 B-02에 비해서 의미 구조가 더 복잡하게 형성되고 있다. ‘화폐’와 ‘물가’가 양쪽에서 중심적인 연결 관계를 형성

하고 있지만 ‘증가’, ‘인플레이션’ 등이 중간에서 다층적인 연결을 하고 있다.



〈그림 4〉 B-03 텍스트의 네트워크 구조



〈그림 5〉 B-04의 네트워크 구조

〈그림 5〉는 일반인을 대상으로 하는 B-04 텍스트의 네트워크 구조를 시각화한 것이다. 〈그림 5〉를 통해 보면, B-04 텍스트는 의미 구조가 분절적으로 나타나는 것이 아니라 구조적인 연결성이 보다 강화된 것을 확인할 수 있다.

‘가격’과 ‘변동’, ‘가중치’ 등의 단어가 중심성이 높긴 하지만 이 외에도 ‘대상’, ‘조사’ 등의 단어가 다층적으로 연결되기 때문에 텍스트의 의미망이 앞의 텍스트들과는 다르게 분절적으로 나뉘지 않고 전체적으로 유기적으로 연결된다.

이상의 분석 결과를 바탕으로 다음과 같은 시사점을 얻을 수 있다.

첫째, 텍스트의 난도가 높아질수록 의미 구조가 다층적으로 연결된다. 특정한 어휘의 중심성이 낮아지며 중심 어휘가 다른 어휘와 연결되는 강도가 증가한다. 이는 앞에서 예로 든 B〈그림 1〉~〈그림 5〉의 변화에서 어렵지 않게 확인할 수 있다. 뿐만 아니라 〈표 4〉를 전체적으로 제시한 그림에서도 이러한 경향성을 발견할 수 있다. A군 텍스트의 경우는 분절적인 구조에서 연결성이 강화되는 구조로 나타났으며, C군과 D군의 경우에는 중심 기능을 하는 어휘의 수가 증가하는 것을 확인할 수 있다.

둘째, 텍스트의 난도가 높아질수록 의미군의 구분이 느슨해진다. 이는 첫 번째 성질과도 관련이 되는데, 텍스트의 난도가 높아지면서 분명하게 나뉘지는 의미군의 형성이 느슨해지고 전체적으로 연결되는 긴밀성이 높아진다. 또한 연결되는 단어 사이의 간격도 넓어지면서 전체적으로 네트워크의 면적이 확장되는 양상으로 나타난다. 이는 단어 간의 의미가 보다 다층적으로 연결되면서 의미의 폐쇄성이 감소하는 것으로 해석할 수 있다. 이러한 전체적 긴밀성은 텍스트의 유기성과도 관련이 있을 것으로 판단된다.

셋째, 어휘의 출현 빈도와 네트워크 중심성은 상호 비례 관계에 있지 않다. 〈표 3〉에서 제시된 최빈 어휘들이 〈표 4〉의 네트워크 구조에서 중심어 역할을 하고 있지는 않기 때문에 핵심어가 반드시 텍스트에 자주 출현하는 단어라고 볼 수는 없다. B-01 ~ B-04 텍스트에서 각 텍스트의 핵심어와 상관관계가 높은 단어들을 추출하면 다음과 같은데, 이는 〈표 3〉에서 제시한 최빈 단어와는 차이가 크다. 또한 전문인을 대상으로 하는 최고 난도의 텍스트에서는 초·중·고등학교급 텍스트와는 달리 상관계수 0.6 이상인 단어의 수가 비약적으로 상승한 것을 알 수 있다.



	B-01 (지폐)		B-02 (지수물가)		B-03 (물가상승)		B-04 (소비자물가지수)	
핵심어	지폐		물가		물가		물가	
순위	단어	상관계수	단어	상관계수	단어	상관계수	단어	상관계수
1	우리	0.87	지수	0.98	상승	0.77	소비자	0.99
2	정도	0.87	느끼다	0.89	경제	0.62	지수	0.97
3	나라	0.76	체감	0.89	공급	0.62	가격	0.92
4	거리	0.66	생각	0.80	기업	0.62	나타나다	0.79
5	늘어놓다	0.66	구매	0.79	생산	0.62	간략	0.78
...								
상관계수 0.6 이상 단어 수	14		12		15		37	

〈표 6〉 텍스트별 핵심어와 상관관계가 높은 단어 목록

## 5. 맺음말

본 연구는 텍스트의 난이도에 따라서 텍스트의 핵심 어휘와 그 핵심어휘들의 관계망이 달라질 것이라는 전제 하에, 텍스트 마이닝 기법을 활용하여 그 양상을 살펴보았다. 기존의 텍스트 이독성 공식과는 다른 방법으로 텍스트의 난이도 차이를 살펴보기 위해, 텍스트의 난이도가 사전에 설정되는 국가수준 학업성취도평가의 읽기 지문을 분석 대상으로 선정하였다. 또한 비교를 위해 일반인 또는 전문인을 대상으로 하는 난도가 높은 텍스트도 선정하였다. 분석은 R 프로그램의 NLP4kec 패키지를 사용하여 텍스트에서 핵심 어휘들을 추출하고 각 어휘들 간의 상관관계를 네트워크 구조로 시각화하였다.

분석 결과 텍스트의 난도가 높아질수록 반복되는 어휘들의 빈도는 낮아지며 반대로 중심 어휘들의 관계성 및 어휘간의 관계성은 증가한다는 것을 확인할 수 있었다. 난도가 높아질수록 의미군이 뚜렷하게 나누어지는 현상이 희석되고 핵심어들 사이의 연결 관계가 다층적으로 형성되었다. 또한 난도가 높아질수록 네트워크의 범위가 넓어지고 간 단어들 사이의 거리도 증가했는데 이는 의미 연결의 다층성을 보여준다고 할 수 있다. 또한 전문적인 글의

경우에는 텍스트의 핵심어와 유의미하게 연결되는 단어의 수가 비약적으로 증가하기도 하였다.

물론 이 연구는 엄밀한 기준으로 텍스트를 선정하고 이를 바탕으로 결론을 일반화하는 양적 연구 방법을 적용한 것은 아니다. 따라서 이 연구는 다음과 같은 한계점을 갖는다. 첫째, 이 연구에서 내린 결론은 분석 대상이 되는 텍스트에 한정된 것일 뿐 이를 난이도에 따른 텍스트의 보편적인 경향으로 해석하는 데에는 무리가 있다. 특히 보편적인 난이도에 따른 결론이라기보다는 텍스트 자체의 영향을 크게 받을 수 있으므로 향후 텍스트의 난이도에 대한 보다 면밀한 구분이 이루어진 이후에 본 연구 방법을 적용할 필요가 있을 것으로 판단된다. 둘째, 텍스트 마이닝기법을 적용하기 전에 해결해야 할 불용어 등의 전처리 문제를 해결해야 한다. <그림 2>와 <그림 4>에서 모두 ‘되다’가 중요한 역할을 하는 것으로 나타나는데 이러한 서술어의 경우 전처리 과정에서 삭제가 가능하다. 이처럼 서술어에서 특정한 의미 기능을 수행하지 않는 어휘 목록을 정리할 필요가 있다. 의존명사는 처리하기가 더욱 까다롭다. 보통 의존명사는 한 글자 단어의 삭제 과정에서 자연스럽게 정리가 되는데, 앞에서도 언급한 것처럼 한국어 텍스트에서는 한 글자의 단어를 함부로 삭제하기가 어렵다. 그 결과 <그림 3>에서는 ‘수’가 중요한 연결 기능을 수행하는 것처럼 그려졌는데 이러한 의존명사도 전처리 과정에서 정리해야 하므로, 향후 한국어 불용어 목록에 대한 추가 연구가 필요하다. 이 외에도 네트워크 구조에서 방향성을 고려해야 하는지 여부, 중심성 지수의 선택, 시각적 구조도의 해석 방법 등 추가적인 후속 연구가 필요하다.

그럼에도 불구하고 이 연구는 텍스트의 구조를 텍스트 마이닝 기법을 사용함으로써 시각적으로 그려낼 수 있는 방법을 시험적으로 제시하였다는 점에서 의의를 갖는다. 또한 기존에 텍스트 빈도 추출에 집중하였던 텍스트 마이닝의 연구 방법을 텍스트의 구조에까지 확장할 수 있는 가능성을 제시했다는 점에서도 연구의 의미를 찾을 수 있다. 이러한 기법은 텍스트에 대한 통계 처리를 바탕으로 하므로 이 과정에서 산출되는 다양한 수치 자료들은 향후 텍

스트 분석 및 평가에 유의미하게 사용될 것으로 판단된다. 이와 같은 연구 성과가 축적되면, 교과서를 제작하거나 평가 지문을 생성하는 장면에서 텍스트를 선정하고 수정하는 데 참고할 수 있는 객관적인 지표로 활용이 가능할 것이다. 또한 한국어 교육에서도 학습자에 따른 텍스트 난이도를 산출하고 교육용 자료를 구성하는 데에도 활용이 가능할 수 있다.



## 참고 문헌

- 강범일·박지홍(2013), 「프로파일링 분석과 동시출현 단어 분석을 이용한 한국어교육학의 정체성 분석」, 정보관리학회지 Vol.30 No.4, 한국정보관리학회.
- 김동성·김종우(2013), 「클라우드 컴퓨팅 관련 논문의 서지정보 및 인용정보를 활용한 연구 동향 분석」, 한국지능정보시스템학회 학술대회논문집 Vol.2013 No.11, 한국지능정보시스템학회.
- 김선아 외(2015), 「21세기 융복합 교육을 위한 미술과·국어과 교육과정 분석 연구: 텍스트 마이닝 기법을 중심으로」, 미술교육논총 제29권, 한국미술교육학회.
- 김용학(2003), 『사회 연결망 분석』, 박영사.
- 노형남(2014), 「빅 데이터 텍스트 마이닝 - 정치 연설을 중심으로 -」, 화법연구 제 26 권, 한국화법학회.
- 박지성·이준석(2017), 「텍스트 네트워크 분석을 위한 대규모 텍스트의 자료정리 방법에 대한 연구」, 현대사회와 행정 제27권 제4호, 한국국정관리학회.
- 이슬기·박영민(2017), 「쓰기 수행 수준에 따른 중학생 논설문의 텍스트 시각화 분석」, 학습자중심 교과교육 연구 제 17권 제15호, 학습자중심 교과교육학회.
- 서혁(2011), 「읽기(독서) 교육 체계화를 위한 텍스트 복잡도(text complexity) 상세화 연구(1) - 텍스트 복잡도 연구의 비판적 고찰을 중심으로」, 국어교육학연구 제 42권, 국어교육학회.
- 서혁 외(2012), 「읽기(독서) 교육 체계화를 위한 텍스트 복잡도 상세화 연구(2)」, 국어교육학연구 제47권, 국어교육학회.
- 성태제(2014), 『현대 기초통계학 - 이해와 적용』, 학지사.
- 윤창욱(2006), 「비문학 지문 이독성 공식 개발에 관한 연구」, 한국교원대학교 교육대학원 석사학위논문.
- 이나래(2010), 「대학수학능력시험 언어영역 비문학 지문의 통사론적 연구」, 한국외국어대학교 석사학위논문.
- 이수경 외(2011), 「한국 간호학 연구주제의 사회 연결망 분석」, Journal of Korean Academy of Nursing Vol.41 No.5, 한국간호과학회.
- 장미경(2012), 「한국어 읽기 교육을 위한 텍스트 난이도 평가 방안 연구」, 고려대학교 박사학위 논문.
- 정옥년(2003), 「읽기 지도에서 수준별 텍스트 활용」, 독서연구 제9권, 한국독서학회.
- 최숙기(2012), 「텍스트 복잡도를 활용한 읽기 교육용 제재의 정합성 평가 모형 개발

연구」, 한국교육과정평가원.

최천택(1995), 「Readability : Theory and Application」, 한신논문집 제12권.

C. Haythrnthwaite(1996), 「Social Network Analysis: An Approach and Technique for the Study of Information Exchange」, Library research, pp.323-342

X. Li, H. Chen, Z. Huang, M. C Roco, 「Patent citation network in nanotechnology」, Journal of Nanoparticle Research Vol.9 No.3, Springer Science, pp.337-352

※ 자료 출처

2009 국가수준 학업성취도 평가 연구 -국어-, 연구보고 RRE 2010-6-2, 한국교육과정평가원.

2010 국가수준 학업성취도 평가 결과 분석 -국어-, 연구보고 RRE 2011-03-02, 한국교육과정평가원.

2011 국가수준 학업성취도 평가 결과 분석 -국어-, 연구보고 RRE 2012-02-02, 한국교육과정평가원.

2012 국가수준 학업성취도 평가 결과 분석 -국어-, 연구보고 ORM 2013-37-01, 한국교육과정평가원.

2013 국가수준 학업성취도 평가 결과 분석 -국어-, 연구보고 ORM 2014-30-01, 한국교육과정평가원.

2014 국가수준 학업성취도 평가 결과 분석 -국어-, 연구보고 ORM 2015-45-01, 한국교육과정평가원.

B\_04\_money :  
[http://www.riss.kr/search/detail/DetailView.do?p\\_mat\\_type=be54d9b8bc7cdb09&control\\_no=af2e165c9c4609caff0bdc3ef48d419](http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cdb09&control_no=af2e165c9c4609caff0bdc3ef48d419)

C\_04\_festival : <http://captaincool.tistory.com/72>

D\_04\_non : [http://news.chosun.com/site/data/html\\_dir/2018/03/23/2018032303095.html](http://news.chosun.com/site/data/html_dir/2018/03/23/2018032303095.html)

이삼형

한양대학교 사범대학 국어교육과  
서울특별시 성동구 왕십리로 222 (사근동)

02-2220-1136

hyung@hanyang.ac.kr

길호현

서원대학교 사범대학 국어교육과  
충청북도 청주시 서원구 무심서로 377-3(모충동)

043-299-8311

kilhohyun@seowon.ac.kr

접수 일자: 3월 27일

심사(수정, 재심) 일자: 4월 1일 ~ 4월 10일

게재 확정 일자: 4월 16일

<부록>

A-01 민화(2012년 초등)

민화는 서민들 사이에서 유행한 그림이다. 민화는 전문 화가가 아니어도 누구나 그릴 수 있었고, 특정한 형식에 얽매이지 않았다. 민화에는 다양한 동식물이 소재로 사용되었는데, 서민들은 이러한 동식물을 청색, 백색, 적색, 흑색, 황색의 화려한 색으로 표현하였다.

민화에는 서민들의 소망이 담겨 있다. 서민들은 민화를 통하여 부귀, 화목, 장수를 빌었다. 예를 들어 부귀를 바랄 때에는 활짝 핀 만드라미나 잉어를 그렸다. 화목을 바랄 때에는 어미 새와 여러 마리의 새끼 새가 함께 있는 모습을 그렸다. 또 장수를 바랄 때에는 바위나 거북 등을 그렸다.

민화에는 나쁜 기운을 물리치고자 하는 서민들의 바람도 담겨 있다. 나쁜 귀신을 쫓아내고 사악한 것을 물리치기 위해 해태, 닭, 개 등을 그렸다. 불이 나지 않기를 바라는 마음에서 전설의 동물 해태를 그려 부엌에 걸었다. 또 여독을 밝히고 잡귀를 쫓아내기 위해 닭을 그려 문에 걸었다. 도둑이 들지를 않기를 바라는 마음에서 개를 그려 공간에 걸었다.

우리는 민화를 통해 서민들의 소망과 뜻을 엿볼 수 있다. 민화에는 현실에서 이루고 싶은 서민들의 소망이 솔직하고 소박하게 표현되어 있다. 또 신비스러운 용을 할아버지처럼 그리거나 호랑이를 바보스럽게 표현하여 재미와 웃음을 찾고자 했던 서민들의 멋스러움도 잘 드러난다. 그림에 담겨 있는 서민들의 소망과 뜻을 찾아 가며 민화를 감상해 보자.

A-02 초상화(2011년 중등)

미술에서 ‘프로파일(profile)’은 사람의 측면을 묘사함으로써 인물의 핵심적인 특징을 뽑아낸 그림을 가리킨다. 서양에서는 중세 말에서 르네상스 무렵에 이런 프로파일 초상화가 많이 그려졌다. 그에 비해 우리나라를 비롯한 동양에서는 프로파일 초상화가 거의 발달하지 않았다. 동양, 특히 중국에서는 오히려 정면상이 발달하였다. 대상의 인품과 특징을 압축적으로 나타내기에 정면상이 더 적합하다고 여겼기 때문이다. 서양에서도 정면상을 그렸지만 그 빈도가 동양보다 낮다.

측면과 정면 중 인물의 특징을 더 잘 나타내는 것은 어느 쪽일까? 우선 동물들의 이미지를 떠올려 보자. 동물들을 그릴 때 정면, 측면, 윗면 가운데 어느 면이 제일 먼저 떠오르는가? 먼저 말을 그려 보자. 말은 일반적으로 옆에서 본 이미지가 가장 먼저 떠오른다. 물고기는 어떤가? 그것도 옆에서 본 이미지이다. 도마뱀을 그려 본다면? 위에서 본 이미지가 제일 먼저 떠오를 것이다. 이런 것들이 우리의 머릿속에 각인된 전형적인 이미지 면이다.

그렇다면 사람은 어떤가? 사람은 다른 동물과 달리 두 개의 경쟁적인 이미지면을 동시에 갖고 있다. 고대 이집트의 벽화가 이를 잘 보여 준다. 이집트 벽화 중에 귀족 ‘네바문’을 그린 그림이 있다. 얼굴과 다리는 측면에서 본 모습이고, 가슴과 눈은 정면에서 본 모습을 그린 것이다. 해부학적으로 불가능한 구성 혹은 자세이지만, 이 그림뿐 아니라 고대 이집트 벽화 대부분이 이런 식으로 그려졌다. 이 혼합 형식으로부터 우리가 확인할 수 있는 것은, 인간이 신체 부위에

따라 정면이 먼저 떠오르기도 하고 측면이 먼저 떠오르기도 하는 존재라는 사실이다.

이렇듯 인간이 두 개의 이미지 면을 동시에 갖고 있기 때문에 동서양 모두 두 이미지 면을 한꺼번에 나타내는 ‘부분 측면상’을 발달시켰다. 부분 측면상은 사람을 완전히 옆에서 보는 것이 아니라 비스듬히 옆에서 보는 것이다. 그러면 정면과 측면의 특징을 동시에 드러낼 수 있다. 그에 비해 고대 이집트 벽화는 인간의 두 이미지 면을 동시에 나타내기 위해 정면과 측면을 신체 부위에 따라 편의적으로 봉합하는 혼합 형식을 이용했다는 점이 흥미롭다.

#### A-03 추상화(2013년 고등)

추상 활동이란 어떤 대상의 특성이나 속성 따위를 추출하여 파악하는 고차원적인 작용이다. 그런데 대상의 새로운 측면을 드러내는 추상 활동은 대상을 보는 관점에 따라 헤아릴 수 없이 많이 존재한다. 이러한 추상 활동 하나 하나는 숨은 진실을 비춰주게 된다. 일반적으로 추상 활동의 결과물은 관찰과 단순화의 방식을 통해 대상의 특징을 포착하는 데서 얻어진다.

피카소의 창작 과정은 추상 활동을 이해하는 데 중요한 단서를 제공한다. 피카소는 ‘황소’ 연작물을 시작하면서 처음에는 황소의 모습을 아주 사실적으로 묘사했다. 그는 황소의 몸에서 평평해진 부분들이 형태를 만들고 있다는 점에 초점을 맞추면서 연작을 그려 나갔다. 그러다가 평면들의 가장자리와 모서리에서 황소의 특징이 잘 드러난다는 것을 깨닫고 간단한 외곽선 몇 개로 황소를 그렸다. 그리고 마지막에 가서는 황소의 몸을 이루는 요소들을 대부분 제거하고 머리의 특징을 잡아내어 그렸다. 이 그림은 몸을 이루는 요소가 사라졌음에도 ‘황소다움’의 본질을 잘 보여 주었다. 피카소에게 황소다움이란 몸의 형태나 크기에 있는 것이 아니라 뿔처럼 아주 단순한 것에 깃들어 있었던 것이다.

이처럼 피카소는 대상을 관찰하는 것이 어떤 예술가에게든 가장 우선적으로 해야 하는 중요한 일임을 강조했다. 피카소는 많은 사람들이 저지르는 실수 중 하나로, 현실을 무시하면서 추상화(抽象化)를 그리는 것에 대해 언급한 바 있다. 그는 추상화를 그리려는 다른 화가들에게 ‘추상에 도달하기 위해서는 항상 구체적인 실재(實在)로부터 시작해야 한다. 무엇인가 실체가 있는 것에서 출발해야만 나중에 실재의 흔적들을 제거해 나가더라도 그 오브제가 표방하는 이념이 사라지지 않는다.’라고 조언했다.

#### A-04 정물화

정물화에 해당하는 프랑스 용어인 ‘nature morte’를 문자 그대로 번역할 경우 죽은 자연이 된다. 죽은 자연이란 물론 존재하지 않는다. 이와 같이 경멸적이고 가혹한 용어는 17세기 프랑스에서 회화를 그 주제에 따라 등급을 매기고 성격을 규정할 때 붙여진 것이다. 그러나 정물화는 결코 죽었다고는 할 수 없는 자연을 대상으로 한 것이며 마찬가지로 결코 자연스럽다고 할 수 없는 죽음을 드러내곤 했다. 정물화를 주제에 따른 엄격한 위계질서 내에서 규정하는 것은 이 세계를 대상으로 하는 모든 회화가 지니고 있는 복잡한 관념 고리 및 이와 같은 개념



연관에서 파생되는 도 다른 의미들을 간과하는 것이다.

예를 들어, 반 고흐의 정물화는 복잡한 암시들로 가득 차 있으며 계속해서 많은 의미들을 파생시키고 있다. 화판 위의 정물들은 아무렇게나 놓인 것처럼 보이지만 사실은 엄격한 규칙 아래 배열되어 있다. 화판의 위쪽 모서리는 작품의 위쪽 모서리와 거의 평행되게 놓여져 있는데 안정감보다 오히려 어떤 불안함을 느끼게 한다. 두껍게 바른 물감 위에 붓질을 하여 목재화판의 질감이 강조되고 있다. 화판 위에 놓인 정물들은 일반적인 정물화에서 흔히 볼 수 있는 것들이다. 주전자, 병, 양초, 책, 야채 등은 아마도 긴긴 정물화의 역사에서 가장 자주 등장하는 소재들일 것이다. 그러나 이 작품의 경우 이들은 보다 적극적인 의미를 지니고 작품을 구성하는 요소이며 결코 작가가 우연히 선택한 것이 아니다.

고호는 죽기 한해 전인 1898년 1월 이 작품을 그렸다. 아를르에서 자신의 귀를 자르는 처참한 사건 이후 동료였던 고갱과 헤어지고 한달 후에 그런 작품으로, 완성된 것으로는 사건 이후 최초의 작품이라고 할 수 있다. 그를 치료하던 의사는 그에게 정신치료요법으로 그림을 그리는 것을 당부했고 고호는 한 달 후 2월, 병이 재발하기 전까지 자신의 예술적 양식과 화가로서의 사명감에 확신을 되찾았던 것처럼 보인다. 이 기간 동안 그는 의사 및 물랭 부인의 초상화와 얼굴을 봉대로 감고 있는 자화상 등 많은 작품을 열정적으로 그려나가고 있었다.

화판 위의 정물은 바로 이러한 상황 속에서 제작된 작품으로 정신적인 안정을 위한 것이다. 음식과 술, 책 그리고 고호에게 언제나 위안을 준 파이프와 담배 등 건강과 안정을 상징하는 정물들이 화판 위에 놓여져 있다. 아마도 파리의 동생으로부터 왔을 편지는 고호와 외부와 연결되어 있음을 의미한다. 책은 건강과 위생학에 대한 책으로 양파처럼 보이는 마늘은 바로 이 책에서 건강식품으로 권장한 것이다. 이 작품 속에서 화판은 작가 자신의 정체성을 회복하기 위해 싸우는 경기장이다. 외형상 작품은 신뢰를 가지고 질서정연하게 구성되어 있지만 과장된 대각선 구도와 겨우 균형을 이루고 있다. 위태로운 긴장이 작품에 감돌고 있으며 대각선으로 불쑥 솟아나와 독단적인 느낌을 주는 촛대와 병이 이러한 긴장을 지탱시켜 준다. 촛대에는 촛불이 켜져 있다. 이 정물화는 고호 자신의 성격을 생생하게 전달해 주는 수단이며 그의 믿음과 희망을 서약하던 장이었다. 고호의 정물화는 인간화되어 있으며 심리학적인 자화상으로 변모하였다.

정물화는 흔히 생각하는 것보다 훨씬 복잡하고 미묘한 의미를 지니고 있다. 물론 그렇다고 해서 모든 정물화에 대해서 고호의 작품처럼 정물화를 그릴 당시 화가의 특정 심리상태에 대해서 우리가 자세히 알아야지만 그 작품을 이해할 수 있다는 의미는 아니다. 정물화는 오히려 전적으로 화가의 테크닉을 과시하기 위해 - 그리고 알면 그 테크닉을 즐기기 위해 - 제작돼 일종의 선전역할을 했던 경우가 더 많았다.

**B-01 지폐(2009년 초등)**

우리나라 지폐의 수명은 다른 나라에 비해 짧다. 그것은 사람들이 지폐를 함부로 다루기 때문이다. 만 원권 지폐의 수명은 4년 반이고, 오천 원권과 천 원권 지폐의 수명은 겨우 2년 정도이다.

우리나라에서 1년에 폐기하는 지폐는 약 10억 3천만 장이다. 이렇게 폐기되는 지폐를 한 줄로 늘어놓으면 서울에서 부산까지 거리의 376배 정도나 된다고 한다. 폐기된 지폐만큼 새로 만드는 데 드는 비용은 약 670억 원이다. 우리 지폐를 함부로 사용하지 않으면 그 비용을 절약할 수 있다. 지폐를 오래 사용하려면 어떻게 해야 할까?

첫째, 지폐에 낙서를 하지 않는다.

둘째, 침이나 음식물 등이 묻지 않게 한다.

셋째, 지갑에 넣어 보관한다.

넷째, 세탁하기 전, 옷에 지폐가 있는지 살펴본다.

다섯째, 지폐를 소중하게 여기는 마음을 가진다.

**B-02 지수물가(2013년 중등)**

지수물가는 가격 변동을 측정하기 위하여 통계적 방법으로 처리된 평균적인 물가이다. 그런데 소비자는 실생활에서 느끼는 체감물가와 통계청에서 발표하는 지수물가가 다르다고 생각한다. 이에 대한 여러 가지 원인 중에서 대표적인 세 가지를 알아보자.

첫째, 지수물가는 대표적인 품목만을 대상으로 한다. 그런데 모든 소비자가 동일한 품목의 물건을 구매하지는 않는다. 그래서 모든 소비자에게 지수물가를 공통적으로 적용할 수는 없다. 중학생이 있는 집에서는 교복, 참고서, 학용품 등의 가격 변화에 민감하지만 중학생이 없는 집에서는 이를 실감할 수 없다. 사람들은 각자가 구입한 물건 값의 변화를 전체 물가의 변화로 생각하는 경향이 있다.

둘째, 지수물가는 전국 주요 도시의 상점과 서비스 업체 중 일부를 표본으로 추출하여 조사한 평균이다. 지수물가가 내려갔다고 할지라도, 개인이 구매한 물건의 가격이 올랐을 경우에 사람들은 물가가 올랐다고 생각한다. 예를 들어, 내가 산 신발 가격이 5만 원이라고 할 때 전국의 신발 평균 가격이 4만 5천원이라고 한다면, 사람들은 자신이 느끼는 체감물가가 지수물가와 다르다고 생각한다.

셋째, 소비자의 기억 차이도 원인이 될 수 있다. 지수물가는 가격이 오른 품목 뿐만 아니라 내린 품목도 대상으로 한다. 그러나 소비자는 가격이 오르고 내린 것에 상관없이 가격이 오른 것만을 오래 기억하는 경향이 있다. 만약 800원 하던 볼펜이 1,000원으로 오르고, 500원 하던 공책이 200원으로 떨어졌더라도 소비자는 가격이 오른 볼펜만을 기억하는 것이다.

통계청이 발표하는 지수물가와 소비자가 느끼는 체감물가의 차이가 커지게 된다면 통계청의 지수물가에 대한 신뢰성이 떨어질 수 있다. 이것을 막기 위해서 많이 구매하는 물건이나 밥상에 자주 오르는 먹을거리를 중심으로 새로운 지수물가를 따로 설정하거나 기준이 되는 품목이나 가중치를 시대의 변화에 따라 바꾸기도 한다.

**B-03 물가상승(2009년 고등)**

2008년 아프리카 남부에 있는 국가인 짐바브웨의 연간 물가 상승률이 2억 퍼센트를 넘었다고 한다. 짐바브웨에서 달걀 3알을 사려면 수레에 돈을 쌓아야 할 형편이었다. 이처럼 화폐가치가 하락하여 물가가 전반적이고 지속적으로 상승하는 경제 현상을 ‘인플레이션’이라 한다. 그렇다면 인플레이션이 일어나는 이유는 무엇일까? 크게 보면 상품에 대한 수요의 용인과 공급의 요인으로 구분해 볼 수 있다.

첫째, 수요 요인에 의한 인플레이션은 대체로 시중 통화량이 늘어서 발생한다. 이를 화폐 수량적 인플레이션이라고 한다. 그런데 늘어난 화폐는 추가되는 수요를 부르고 소비의 증가를 가져온다. 그리고 늘어난 수요에 맞춰 공급자는 가격을 올리게 된다. 이를 수요 견인형 인플레이션이라고 한다.

통화량과 물가 사이의 관계를 이해하는 데 유용한 것이 경제학에서 유명한 어빙 피셔의 ‘화폐 교환 방정식’이다.

$MV=PT$

M: 통화량, V: 화폐가 유통되는 속도(혹은 횟수),

P: 물가 수준, T: 거래량

이 방정식에 의하면, 일정 기간 경제 전체에 거래되는 양(T)과 화폐의 유통 속도(V)가 동일하다고 가정할 때, 통화량(M)이 증가하면 물가 수준(P)가 올라간다. 반면, 시중에 통화량이 늘어도 유통 속도가 준다면 물가가 인상되지 않을 수도 있다. 하지만 보통 통화량이 늘면 화폐의 유통 속도도 증가하기 마련이다. 그래서 소비자의 지출이 증가해도 인플레이션이 일어날 수 있다.

둘째, 소비가 늘어나지 않아도 공급 측면에서 물가가 오를 수도 있다. 제품 생산 비용의 상승으로 인해 제품 가격이 인상되어 물가가 상승하게 되는 것이다. 이를 비용 인상형 인플레이션이라고 한다. 비용 인상형 인플레이션의 원인으로 원자재 가격 상승, 임금 인상, 기업이익(高)이윤 추구 등을 들 수 있다.

캘리포니아의 가뭄으로 산지(產地) 오렌지 값이 올라서 국내 오렌지 주스 값이 오르거나, 국제 유가가 올라서 국내 휘발유 값이 올라가는 현상은 원자재 가격 상승으로 인한 인플레이션이다. 또한 제품 생산의 측면에서 임금은 일종의 비용이므로, 임금이 인상되면 제품의 가격이 오르고 이 때문에 물가가 오른다. 기업이 기술혁신 없이 똑같은 제품에 이윤을 더 붙여 가격을 높여도 물가 상승의 원인이 될 수 있다.

**B-04 소비자물가지수**

([http://www.riss.kr/search/detail/DetailView.do?p\\_mat\\_type=be54d9b8bc7cdb09&control\\_no=af2e165c9c4609caffe0bdc3ef48d419](http://www.riss.kr/search/detail/DetailView.do?p_mat_type=be54d9b8bc7cdb09&control_no=af2e165c9c4609caffe0bdc3ef48d419))

소비자물가지수는 가구가 일상생활을 유지하기 위해 구입하는 각종 소비재나 서비스 가격의 변동을 측정하기 위한 지수로서 상품과 서비스 가격의 변동을 종합적으로 나타내는 지수이다. 소비자물가지수는 가구가 소비하는 재화의 물가를

가중평균한 것으로 각 재화별 가중치가 필요하다.

우리나라의 경우 소비자물가지수의 가중치는 소비자물가 조사 대상 481개 품목의 가격변동을 종합하고자 할 때, 소비생활에 미치는 영향이 품목마다 서로 다르다는 점을 반영하고자 하는 것이다. 동일한 가격상승이라 할지라도 품목에 따라 가계에 미치는 영향은 다르므로 품목별 월평균 소비지출비중을 기초로 하여 품목별 가중치를 작성함으로써 소비지출규모에 비례하는 소비자물가지수를 산정하고자 하였다.

가중치의 모집단은 2010년 가계동향조사 결과 중 의제지출비를 포함한 소비지출액 중 자가보유비와 복권, 중고차 구입비 일부 등을 제외한 금액이다. 소비자물가지수는 대표성을 유지하기 위해 가계동향조사의 월 평균소비지출액의 1/10,000 이상인 481개 품목을 대상으로 산정된다. 또한 소비자물가지수는 이 가중치 모집단(소비지출총액)을 1,000.0으로 하고 개별품목이 차지하는 지출액 비중을 그 품목의 가중치로 사용하여 가격변동을 가중 평균하여 작성된다.

<표 1>은 통계청에서 제공하고 있는 소비자물가지수 데이터를 간략화 하여 나타낸 것이다. 소비자물가지수는 크게 기간, 지역, 품목분류 수준으로 구성된다. 소비자물가지수는 매월 조사 되어 발표 되므로 월별, 분기별, 연도별 자료를 1975년부터 2012년까지의 기간을 대상으로 제공하고 있다. 지역별로는 전 도시 및 37개 시, 9개 구별소비자물가지수를 제공한다. 하지만 지역별로 기준시점의 가격수준이 다르게 때문에 지역별 소비자물가지수를 이용하여 지역별 가격 차이를 비교할 수는 없다. 마지막으로 소비자물가지수는 총 두 개의 품목 분류수준으로 구성된다. 즉, <표 1>과 같이 상위분류인 빵 및 곡물이 있고 이의 하위분류로서 쌀, 잡쌀, 보리쌀 등이 있는 것이다.

#### C-01 지역축제(2009년 초동)

우리나라의 많은 지역에서는 지역 축제가 벌어진다. 지역 축제는 그 고장의 문화를 알리고 발전시키기 위한 것이다. 그런데 우리 고장에는 고장의 문화를 알릴 수 있는 지역 축제가 없다. 우리 고장에서도 지역 축제를 열어야 하는 이유는 무엇일까?

첫째, 우리 고장의 독특한 문화를 널리 알릴 수 있다. 축제를 열면 다른 지역 주민들이 많이 찾아와 우리 고장의 문화를 알게 되기 때문이다. 둘째, 지역 축제를 열면 우리 고장의 경제를 발전시킬 수 있다. 지역 축제를 열면 많은 관광객이 방문하게 되어 경제적인 소득을 얻을 수 있기 때문이다. 셋째, 지역 축제를 열면 우리 고장의 환경오염을 줄일 수 있다. 축제를 열면 사람들이 우리 지역의 주거 환경에 대해 관심을 갖게 되기 때문이다. 넷째, 지역 축제를 계속 열다 보면 주민들의 공동체 의식을 높일 수 있다. 문화는 지역 주민을 하나로 묶어주는 바탕이기 때문이다.

그렇다면 지역 축제를 잘하기 위해서는 어떻게 해야 할까? 구체적인 방법을 정하기 전에 먼저 다른 고장에서 개최되는 축제들의 문제점을 알아볼 필요가 있다. 그렇게 할 때 좀 더 나은 방법을 찾을 수 있기 때문이다.

다른 지역에서 개최하는 축제의 문제점은 <그림1>과 같다. 지역 축제가 지나치게 상업적이고 지역의 고유한 특성을 사리지 못하는 것이다.

C-02 싱싱축제(2010년 중등)

남태평양 남서부에 있는 파푸아뉴기니는 세계에서 두 번째로 큰 섬인 뉴기니의 동쪽 절반을 차지하고 있는 나라이다. 파푸아뉴기니에는 800여 부족이 700여 종의 다양한 언어를 사용하며 살고 있다. 그리고 이들 중 상당수는 아직도 사냥과 채집을 하며 정글 깊은 곳에 흩어져 산다.

각 부족들은 각자 다른 풍습과 문화를 가지고 있기 때문에 서로 갈등을 겪는 경우가 많았다. 파푸아뉴기니 정부는 2년마다 전 부족이 참가하는 대규모 축제를 조직하게 되었는데, 이것이 바로 싱싱 축제이다. 이때 ‘싱싱’이란 전통 춤과 노래를 일컫는 말이다.

이 축제는 독립 기념일인 9월 16일을 전후해 수도인 포트모르즈비에서 열린다. 축제에는 전국의 모든 부족들이 참여하며, 이 기간에 사람들은 모든 일을 접어두고 오직 축제에만 온 정열을 쏟는다.

싱싱 축제의 주요 행사는 싱싱 경연 대회이다. 싱싱 경연 대회에 참가한 여러 부족들은 온갖 치장을 한다. 남자들은 풀과 나무, 꽃에서 추출한 안료를 이용해 얼굴에 강렬한 색을 칠하고, 각종 새들의 깃털로 장식도 한다. 이렇게 화려하게 꾸미고 난 후, 사람들은 전통에 따라 화살이나 나무로 만든 칼을 들고 춤을 추며 그들만의 신을 불러내는 다양한 의식을 행한다. 이 의식에는 다른 부족으로부터 자신의 부족을 지키려고 했던 사람들의 심리가 반영되어 있다.

싱싱 축제의 또 다른 행사는 포트모르즈비의 엘라 해변에서 열리는 ‘히리 모알레’이다. ‘히리 모알레’란 ‘행복한 무역’이라는 뜻으로, 교역을 무사히 마치고 돌아오는 남자들을 환영하기 위해 여자들이 춤추고 노래하던 풍습을 재연한 것이다. 이 행사에서 여자들은 춤과 노래로 남자들 못지않은 열정을 과시한다.

이처럼 다양한 행사로 치러지는 싱싱 축제는 파푸아뉴기니 사람들의 열정으로 가득 차 있다. 이 축제는 각 부족이 그들의 전통과 힘을 과시하는 장(場)인 동시에, 다른 여러 부족들과 하나로 뭉치는 화합의 장이기도 하다.

C-03 문화관광(2009년 고등)

‘관광(觀光)’이라는 말은 원래 중국의 역경(易經)에 나온 “관국지광(觀國之光) 이용빈우왕(利用貧于王)”, 즉 “나라의 빛을 두루 보면 왕의 좋은 신하가 된다.”라는 구절에서 나왔다. 이 때 ‘빛’이란 다른 나라나 지역의 정치적·경제적 제도를 이르는 말이다. 어원을 살펴보면 관광에는 이제껏 보지 못한 제도를 찾아 나선다는 의미가 있었다. 그러나 일반적으로 사람들은 관광이라고 하면 휴양 관광을 먼저 떠올려 왔다. 휴양 관광은 보통 몸과 마음에 쌓인 피로를 해소하고, 실상으로 다시 돌아가서 일하거나 생활할 수 있는 활력소를 얻고자 하는 목적이 있다.

그런데 요즘 경제적 여유와 높은 교육 수준을 갖춘 계층을 중심으로 한 ‘문화관광’이 부각되고 있다. 문화 관광객들은 미리 짜인 일정대로 움직이며 휴양을 즐기는 관광에서는 별다른 감흥을 얻지 못한다. 대신 이들은 자신의 지식과 교양을 토대로, 차별화된 경험이 기대되는 곳을 찾아가 그 관광지만의 독특한 문화를 발현하는 데 열중 한다. 요즘에는 이들의 성향을 맞추어 고고학적 유물이나 옛날 건축물, 예술품, 민속 문화 등 예전에는 인류학자나 고고학자, 문화 사

학자지나 관심을 둘 만한 요소를 테마로 한 관광 프로그램이 발전하고 있다.

그러나 아쉽게도 아직도 많은 관광지들은 관광객들을 대상으로 대중적인 볼거리를 제공해야 한다는 시각에만 갇혀 있다. 그래서 관광지들에 비슷비슷한 내용과 형식의 축제가 만들어 지고 있다. 모든 관광지의 가판대를 장식하고 있는 똑같은 기념품들도 그곳만의 독특한 문화적 가치를 형성하는 데 방해가 되는 대표적인 요인이다. 편리한 교통과 안락한 시설은 관광객을 위해서 필요한 요소이지만, 이것만을 중요시하는 것은 관광에 대한 기존 관점의 한계를 드러낼 뿐이다. 문화 관광객들은 다른 곳에서는 찾아보기 힘든 독특한 문화적 가치를 체험하고 싶은 것이 일상에서 벗어난 휴식이나 편안함을 원하는 것은 아니기 때문이다.

독특한 문화 체험을 바라는 관광객들이 추구하는 바는 어느 하나의 가치로 수렴되지 않는다. 그런 만큼 예전의 휴양 관광 산업에서 요구되는 기준을 그대로 따라갈 수는 없다. 따라서 관광 산업에서 바라보는 시각을 관광객들의 요구에 맞추고, 이들에게 새로운 문화 체험을 할 수 있게 하는 관광 프로그램을 개발해 나가야 할 것이다.

#### C-04 리우 카니발(<http://captaincool.tistory.com/72>)

지상 최대의 축제인 리우 카니발은 세계 3대 미항인 브라질의 리우데자네이루에서 열린다.

매년 2월 말부터 3월 초 사이 4일 동안 개최되며 이 때는 토요일 밤부터 수요일 새벽까지 밤낮을 가리지 않고 축제가 열린다.

리우 카니발은 포르투갈에서 브라질로 건너온 사람들의 사순절 축제와 아프리카 노예들의 전통 타악기 연주와 춤이 합쳐져서 생겨났다. 1930년대 초반까지는 보통의 거리축제에 지나지 않았지만 그 후 삼바학교들이 설립되고 학교별로 퍼레이드를 펼치면서 지금과 같은 큰 규모의 축제로 발전하였다. 삼바학교는 카니발을 위한 학교이다.

사람들은 삼바학교에 등록하여 1년 동안 퍼레이드를 준비한다. 리우 카니발의 핵심은 삼바 퍼레이드이다. 삼바 무용수들이 퍼레이드를 할 수 있도록 설계된 거리를 '삼바드로모'라고 하는데, 총 6만 명을 수용할 수 있다.

또 타악기를 연주하는 대규모 밴드를 '바테리아'라고 한다. 삼바 퍼레이드에서 한 그룹마다 삼바 춤을 추는 사람만 약 4000명에 이른다. 해마다 리우 카니발이 열릴 때면 전 세계에서 약 6만 명의 관광객이 찾아오고, 브라질 국내 관광객도 25만여 명에 이른다.

브라질을 찾는 전체 관광객의 3분의 1에 해당되는 사람들이 리우 카니발이 열리는 시기에 맞춰서 온다고 한다.

**D-01 올바른 비판 문화(2012년 고등)**

“몸에 좋은 약이 입에 쓰다.”라는 말이 있다. 좋은 약은 병을 낫게 하지만 쓴맛 때문에 먹기는 괴롭다. 하지만 입에 쓰다고 약을 주지 않거나 먹지 않으면 병을 고칠 수 없다. 비판도 마찬가지이다. 비판하기를 꺼리거나 비판을 제대로 듣지 않으면 갈등이 심해지고 문제가 더 커질 수 있다. 갈등이나 문제를 해결하여 건 강한 사회를 만들기 위해서는 올바른 비판 문화를 형성할 필요가 있다.

올바른 비판 문화를 형성하기 위해 먼저 비판의 의미를 알아보자. 비판과 비난을 혼동하는 경우가 있는데 비판은 비난과는 엄연히 다르다. 비판은 어떤 행동이나 의견에 대해 이성적으로 판단하여 말하는 것이다, 반면, 비난은 감정만 앞세워서 상대방을 이유 없이 헐뜯는 것이다. 예를 들어, 친구의 글을 읽고 “네 글이 잘 이해되지 않아. 자세한 예를 들어 주면 더 좋을 것 같아.”라고 말하는 것은 비판이라고 할 수 있다. 그러나 “너는 글을 참 못 쓰는 것 같아.”라고만 말하는 것은 비난에 가깝다. 비판은 부족한 점을 흠잡는 것이 아니라 상대방에게 도움을 주는 것이어야 한다.

그렇다면 올바른 비판 문화를 만들어 가기 위해 어떻게 해야 할까? 첫째, 비판을 할 때는 상대방을 존중하는 마음을 가져야 한다. 둘째, 비판을 할 때는 알맞은 이유를 들어야 한다. 셋째, 비판을 할 때는 문제를 해결할 수 있는 대안을 함께 제시해야 한다. 넷째, 비판을 들을 때에는 자신의 생각과 비교하여 받아들여야 한다. 다섯째, 비판을 들을 때에는 상대방의 말을 경청하고 감사의 뜻을 표해야 한다.

비판의 의미를 잘못 이해하여 바르게 비판하지 못하면 상대방에게 상처를 입히거나 상대방으로부터 상처를 받기 쉽다. 따라서 비판의 의미를 제대로 이해하고 올바른 비판 문화를 만들어가야 한다. 좋은 비판은 우리 사회를 성숙하고 아름답게 만드는 디딤돌이 될 것이다.

**D-02 강화갯벌(2010년 중등)**

정부는 최근 ‘강화 갯벌’에 조력 발전소를 세울 계획을 발표하였다. 정부의 계획은 조수 간만의 차가 8m나 되는 이 지역에 설비 용량 132만kW급 규모의 조력 발전소를 건설한다는 것이다. 이것은 현재 세계 최대 규모인 프랑스 ‘랑스 조력 발전소’의 다섯 배가 넘는 규모이다.

물론 갯벌은 그 자체로 가치가 있고, 갯벌이 파괴될 때의 피해도 무시할 수 없는 것이 사실이다. 그러나 이곳에 발전소가 만들어지면 한 해의 24억 1,000만 kWh의 전기를 만들 수 있는데, 이는 인천광역시의 각 가정에서 1년 동안 쓰는 전력의 약 60%를 감당할 수 있는 양이라고 한다. 그 결과 한 해에 354만 배럴의 원유 수입 대체 효과가 생겨 전력 수급 문제를 상당 부분 해결할 것으로 기대된다. 더욱이 공사 기간 동안 약 6만 4,000명의 고용 효과가 발생할 것으로 분석되어, 경제적으로 어려운 시기에 정부 차원에서는 포기하기 어려운 사업이다.

또한 갯벌 개발로 인해 발생하는 환경 문제는 발전소 건설 과정에서 나오는 흙을 이용해 인공 습지를 만들어 해결한 예정이라고 한다. 그리고 저어새를 비롯한 멸종 위기의 동식물은 대체 서식지를 만들어 보호할 수 있다고 하니 발전

소 건설은 여러모로 타당하다. 그러므로 경제적 이익과 장기적인 전력 수급 문제를 생각할 때 조력 발전소를 건설해야 한다.

강화 갯벌에 조력 발전소가 들어선다고 한다. 조력 발전소가 건설되면 녹조 현상이 일어나 생태계가 파괴될 수 있다. 또한, 강화 갯벌은 천연기념물 제419호로 지정될 정도로 가치가 높아 갯벌 개발에 대한 우려의 목소리가 높다.

그렇다면 강화 갯벌이 지니는 가치는 무엇인가? 우선, 강화 갯벌은 자연 정화 가치를 지닌다. 강화 갯벌은 서해안 갯벌과 비무장지대 및 한강이 만나는 곳으로, 온갖 오염 물질이 이 갯벌을 거쳐 서해 먼 바다로 가게 된다. 이때 강화 갯벌에 사는 수많은 미생물은 바다로 나가는 오염 물질을 효과적으로 분해한다. 이를 하수 처리장의 운영 비용으로 환산하면 헥타르당 384만원의 가치를 가진다고 한다.

그리고 생태계가 잘 보존된 갯벌은 관광 자원으로 활용되어 여가 가치를 지니기도 한다. 전라남도 남해안에 있는 순천만의 경우 랍사르 총회 때 국내에서는 처음으로 공식 방문지로 지정되면서 방문객이 눈에 띄게 늘었다. 그 결과 2009년 한 해 동안 얻은 소득만 해도 약 1,000억 원에 이른다고 한다. 강화 갯벌도 아시아 습지 보호 협약에 등록된 세계 주요 습지라는 점을 생각한다면 순천만처럼 관광 자원으로 활용할 수 있을 것이다.

세계적인 과학 잡지 『네이처』에 따르면, 이런 갯벌의 생태적 가치는 헥타르 당 9,900달러라고 한다. 농경지의 92달러보다 100배 이상이라는 평가는 눈여겨볼 만하다. 이처럼 갯벌은 자연 그대로 두어도 가치가 크다. 따라서 강화 갯벌은 보존되어야 한다.

#### D-03 전자민주주의(2010년 고등)

대의제는 근대 사회 및 산업 사회의 전형적인 정치 제도로 자리 잡았다. 하지만 민주주의의 이상을 실현하는 데에는 한계가 있는 것도 사실이다. 국민의 의사를 국민의 입장에서 제대로 대변하지 못한 데 따른 ‘대표성의 위기’와, 국민의 정치 참여가 현저하게 줄어들게 되는 ‘참여의 위기’가 심각한 문제로 대두하게 되었다. 그런데 정보 통신 기술이 눈부시게 발전하면서부터 사정은 크게 변화하기 시작했다. 쌍방향 의사소통을 바탕으로 한 인터넷 등의 뉴미디어의 출현으로 민주주의의 이상을 실현하려는 시도가 나타났는데, 이 시도가 곧 ‘전자민주주의’라고 말할 수 있다.

전자민주주의란 대의제 민주주의의 내적 결함을 극복하고 그 대안을 모색하기 위해 성립된 개념이다. 이를 통해 전자 투표, 일반 주민과 정치·행정 지도자와의 의견 교환, 정치·행정 정보의 원활한 소통, 사회적 의제에 대한 공개, 토론 등이 실현될 수 있다.

전자민주주의에 대해 낙관하는 정보사회론자들은 기술을 사회의 발전과 진보를 창조하는 원인으로 간주하는 기술결정론의 시각을 갖는다. 그들은 뉴미디어를 정치적으로 활용함으로써 시민의 정치 참여의 기회를 증대시키고 기존의 사회적 불평등을 해소할 수 있다고 보며, 정보 불평등 현상은 일시적이라고 여긴다. 그리고 뉴미디어를 정치적으로 활용하여 궁극적으로는 시민참여 정치와 직



접 민주주의를 실현함으로써 민주 정치 발전에 기여할 수 있다고 생각한다.

반면 전자민주주의에 대해 회의적인 정치선택론자들은, 기술이 아무리 발달하더라도 그것이 정치에 미치는 결과는 기술 자체보다는 기술을 사용하는 주체의 의식과 사회 제도적 구조에 더 큰 영향을 받는다고 보는 사회선택론의 시각을 갖는다. 그들은 뉴미디어의 출현으로 이론상으로는 국민들의 정치의식이 제고될 수는 있으나, 방관자적 참여 내지 비활동적 참여를 촉진시킬 수 있다고 본다. 이 관점에서는, 인터넷 공간의 수많은 정보 속에서 실제로 개별 이용자가 관심을 갖는 부분은 대단히 제한적이므로 자신의 구체적인 관심 사안에 대해서만 정보를 취하거나 논의에 참여하는 경향이 있다고 본다. 그래서 공공 영역에 대한 관심이 오히려 줄어들 수 있다고 생각한다. 그리고 정보 통신 기술의 발달은 정보라는 자원을 중심으로 또 다른 불평등의 구조를 더하여 사회적 불평등을 심화시킬 수 있다고 여긴다. 아울러 권력에 의한 사회적 감시와 통제의 범위나 방법이 증대할 수 있음을 조심스럽게 지적하고 있다.

정보사회론자들의 주장처럼 기술적 차원에서의 정보화는 현대 민주주의가 안고 있는 많은 문제점을 해결해 줄 수 있는 가능성을 분명히 제공해 준다. 그러나 정치선택론자들이 우려하듯, 전자민주주의의 미래가 결코 희망적이며 낙관적인 것만은 아니다. 전자민주주의에 함축되어 있는 민주주의의 가치를 실현하기 위해서는 정보의 생산·소비 과정에서 민주성과 합리성을 확보하는 일이 선결 과제이다. 이러한 의미에서 전자민주주의의 실현 가능성을 확보하기 위해서는 정보를 생산·소비하는 주체들의 합리적인 인식과 태도를 통해 바람직한 정치 환경을 새로이 설정해 나가는 노력이 요구된다.

#### D-04

통상전쟁([http://news.chosun.com/site/data/html\\_dir/2018/03/23/2018032303095.html](http://news.chosun.com/site/data/html_dir/2018/03/23/2018032303095.html))

트럼프 미국 대통령이 중국산 제품에 최대 600억달러 관세를 부과하는 행정명령에 서명해 대(對)중국 통상전쟁의 포문을 열었다. 이름부터 '중국의 경제침략'을 표적으로 하는 행정명령'이다. 정상적 국가 관계에선 쓰지 않을 '경제침략'이란 극단적 용어까지 붙였다. 중국은 즉각 미국산 농산물 등에 30억달러 맞보복 관세 부과 조치를 발표했다. "끝까지 싸우겠다"는 '항전(抗戰)' 성명도 나왔다. 세계 1·2위 경제 대국이 마치 교전이라도 벌이듯 험한 말을 퍼부으며 정면 충돌로 치달았다. 그 충격으로 한국 증시가 3~4% 폭락하는 등 글로벌 금융시장이 출렁였다.

이날 두 나라가 서로에게 쏘아댄 '600억달러' 대 '30억달러'의 관세 폭탄은 이제부터 본격화될 통상전쟁의 시작일 뿐이다. 트럼프는 미국 IT 기업 등이 중국에 투자하는 것을 제한하고 중국의 지식재산권 침해를 세계무역기구(WTO)에 제소할 것도 지시했다. 중국도 미국산 항공기 등을 맞보복 대상에 추가할 가능성을 밝히고 있다. 미국의 금융 시스템에 타격 주려 보유 미 국채를 매각할 수 있다는 얘기도 흘러난다. 충돌이 제조업 관세에 이어 기업 투자와 무형자산, 금융 등의 분야로 확산될 조짐을 보이고 있다.

이 배경엔 미국이 중국을 바라보는 전략적 인식의 근본적 변화가 자리 잡고 있다. 지금까지 미국은 중국을 자유무역 질서에 편입시켜 정치적 민주화를 유도하는 것이 미국의 이익이라고 보았다. 그래서 천문학적 무역 적자도 감수했다. 그러나 중국이 자유무역의 막대한 이익을 누리면서도 미국과 경쟁하는 정치·군사 패권국가의 길을 질주하자 미국의 통상 전략도 공세로 전환했다. 절대 권력 독재자가 된 중국의 시진핑 주석은 '중화(中華)'를 내세우며 미국 주도의 질서를 인정하지 않겠다는 뜻을 분명히 하고 있다. 통상 격돌이 일시적 현상이 아니라 지속적으로 이어질 구조적 문제가 됐다는 뜻이다.

미·중 충돌로 세계에서 가장 큰 피해를 입을 나라를 꼽으라면 한국이 될 수밖에 없다. 중국·미국은 우리의 1·2위 무역 상대국이다. 우리 수출에서 중국 시장은 25%, 미국은 12%를 차지한다. 미·중 마찰로 무역이 위축되면 수출 의존도가 높은 한국 경제에 큰 타격이다. 특히 중국에 대한 부품·중간재 수출이 직격탄을 맞게 된다. 나아가 미·중 간 통상 패권 경쟁에서 미국이나, 중국이나의 선택을 강요받는 전략적 딜레마에 빠질 수 있다. 트럼프는 '반중(反中) 통상 동맹'을 추진하면서 우방국 참여를 요구하고 있다. 곧 우리에게도 압박이 올 것이다. 지금까지 우리는 '안보는 미국, 경제는 중국'의 분리 노선을 취해 왔지만 이제 한계에 다다르고 있다.

숨 가쁘게 몰아치는 통상 태풍 앞에서 정부 대응은 미덥지 않다. 통상 업무를 산업부 내 차관급 조직에 불과한 통상교섭본부에 맡겨놓고 정부 전체는 뒷짐을 진 것 같은 모습이다. 트럼프는 안보 위협을 이유로 무역 보복을 가하는데 정부는 "안보와 통상은 별개"라는 입장이다. 수단도 힘도 없으면서도 대통령은 "당당하고 결연한" 대응을 주문했다.

한국과 같은 통상 국가에서 글로벌 차원의 무역 전쟁은 국가 생존에 직결되는 문제다. 우리가 가진 모든 역량과 자산을 총동원해 대응하지 않으면 국가적 재앙이 될 수 있다. 청와대를 컨트롤 타워로 삼아 통상·외교안보·국방·정무 라인을 망라하는 범정부, 범국가 차원의 대응 체계를 갖춰야 한다. G20 같은 다자간 대화 채널을 통해 미·중으로부터의 양자(兩者)적 직접 압력을 피하고, 우리와 이익을 같이하는 나라들과 연합 전선을 펴는 방안도 필요하다. 북핵 문제와 동시에 터진 통상 악재여서 위기감을 더 높인다. 모든 대응의 지휘봉은 대통령이 직접 잡을 수밖에 없다

<Abstract>

## A Study on the Changes in the Core Terminology and Related Network of Text

Lee Sam-Hyung(Hanyang Univ.) · Kil Ho-Hyun(Seowon Univ.)

This study used text mining techniques to analyze the relevant network of key words according to their difficulty level. The analysis targets were selected as reading fingerprints on the NAEA(National Accessment of Educational Achievement). It also selects difficult texts targeting professionals. The analysis used the NLP4kec package of the R program. And the correlation between each vocabulary was visualised in the network structure

The analysis showed that as the degree of text increases, the frequency of repeated vocabulary increases, while the relationship between the main vocabulary increases. As the text became more difficult, the network became larger and the distance between the words increased. In the case of professional writing, the number of words associated with the keywords in the text also drastically increased.

This study holds significance in that it has experimentally proposed a way to visualize the structure of text by using text mining techniques. It is also significant that the study showed the possibility of extending the study methods of text mining, formerly focused on extracting text frequencies, to the structure of text.

However, this research is limited to the fact that it is not intended to select text on a strict basis and to generalize conclusions based on it.

\* Key words : text difficulty, core words, key words, text mining, R, SNA(Social Network Analysis)