

# 텍스트 마이닝을 활용한 한국어 학습자의 핵심 어휘 및 문법 사용 양상 연구\*

- 내러티브 텍스트를 중심으로 -

채윤미·길호현·왕연\*\*

## <차 례>

1. 서론
2. 선행 연구
3. 연구 방법
4. 내러티브 텍스트의 구조별 분석 결과
5. 결론 및 제언

## 1. 서론

내러티브(narrative)란 사람이 언어를 통해 사회적 관계를 맺고 삶을 영위하는 데 있어서 매우 중요한 위치를 차지한다. 만난 사람에게 자신을

---

\* 이 논문은 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2018S1A5B5A01034698).

\*\* 채윤미: 한양대학교 국어교육과 조교수, 제1저자.

길호현: 공주대학교 국어교육과 조교수, 교신저자.

왕연: 한양대학교 박사 수료, 공동저자.

소개하는 일이나 친구와 다투게 된 경과에 대한 설명, 할머니께 들은 옛날이야기, 수업 시간에 들은 내용을 친구에게 전달해 주는 것까지 내러티브는 우리 삶에 매우 다양한 형태로 실현된다.

내러티브는 크게 내용을 꾸며 쓰는 허구적 내러티브(fictional narrative)와 실제 경험을 토대로 하는 개인적 내러티브(personal narrative)로 구분되며, 실제 경험을 토대로 하는 개인적 내러티브는 제2언어 학습에 유용하게 적용된다. 실제로 학습 자료의 구어 담화, 문어 텍스트의 상당 부분이 내러티브로 구성되어 있으며 이는 학습자가 수행해야 하는 과제로도 연결된다. 특히 쓰기 영역에서의 내러티브는 개인적 경험이 누구나 소유하고 있는 제재라는 점에서 다양한 글쓰기 활동에 노출되지 않은 학습자라도 어렵지 않게 접근이 가능하고, 아이디어 생성이 큰 문제로 간여되지 않아 표현 전략이나 기술을 학습하는 데 용이하기 때문에 학습자가 목표 언어를 학습하는 초기부터 접하게 된다.

한국어 쓰기 교육 현장에서도 일기나 기행문과 같이 정서나 감정, 경험, 일상생활 등에 대해 기술하는 과제가 초·중급 단계에서 높은 비중으로 다루어지며, 나아가 고급 단계에서는 자전적 소설 쓰기, 문화 유적 답사와 같은 고차원의 내러티브 생산까지 학습 목표로 삼고 있다. 이렇듯 내러티브 텍스트에 대한 쓰기 교육은 필수적이며 매우 중요하지만, 실제 이에 대한 연구는 미비하다. 한국어 학습자를 대상으로 하는 장르 기반 쓰기에 관한 연구는 주로 특수 목적 한국어 쓰기 교육의 일환으로 이루어져 왔으며 특히 논문, 업무 이메일, 보고서 등과 같이 장르성이 비교적 뚜렷한 텍스트에 대해 치우쳐 있는 경향을 보인다. 이에 반해 내러티브 텍스트에 대한 연구가 미진한 이유는 내러티브가 갖고 있는 일상성과 편재성, 규정하기 힘든 무한한 형식 때문일 것이다. 단순히 과거의 일을 기술하는 등의 개인적 경험부터 직접 경험하지는 않았지만 자신이 보고 들은 사건을 전달하는 비개인적 내용에 대한 서술까지도 내러티브에 해당되기 때문에 연구의 대상과 범위를 정하기에 어려운 점이 있다. 이는 교육 현장에도 그대로 대입이 되는데, 논문과 같은 학술적 텍스트나 논증

텍스트의 경우 텍스트의 구조와 언어적 특성에 대한 연구 결과가 그간 꾸준히 축적되어 교육 내용으로 적용되는 경우가 늘고 있는 반면, 내러티브 텍스트는 여전히 학습자가 가지고 있는 이야기 구성 능력에 의존한 채 성공적인 쓰기 과제의 수행을 온전히 학습자의 몫으로만 남기는 경우가 많다. 이러한 까닭에 학습자의 숙달도가 높아져도 내러티브 텍스트의 질적 수준은 향상되지 않는 경우를 적지 않게 볼 수 있으며, 이에 대한 진단이나 평가 기준, 교육 내용도 미흡한 실정이다.

내러티브가 필연적으로 갖게 되는 삶의 경험을 이야기하는 장르라는 점에서 한국어 교육의 전반적인 단계에서 교육이 이루어져야 하며, 언어적 특성에 관한 연구 역시 다양한 각도에서 진행될 필요가 있다. 아울러 실제성 있는 자료를 기반으로 모여 화자와의 비교 분석이 이루어진다면 이는 교육 내용 마련에 도움이 될 수 있다. 또한 이러한 연구가 계량적이고 객관적인 측면에서 이루어질 수 있다면, 그 전형성을 파악할 수 있을 것이다.

이에 본 연구는 빅데이터 분석에서 활용되는 텍스트 마이닝과 언어 네트워크 분석 기법을 활용하여 한국어인 모어 화자와 외국인 한국어 학습자의 내러티브 텍스트에 나타난 어휘와 문법의 사용 양상을 비교해 살펴보고자 한다. 먼저 핵심 어휘를 추출한 뒤 어휘들 간의 관계성을 네트워크를 통해 알아볼 것이다. 텍스트에서 핵심 어휘와 연계 어휘는 그 의미를 통해 해당되는 사고 기능을 표상할 수 있기 때문이다.<sup>1)</sup> 필자는 실제 문장이나 문단을 완성하기 전에 쓰고자 하는 대상에 대해 추상적인 아이디어를 구체화하는데, 이 과정에서 글의 뼈대를 이루는 핵심 어휘 혹은 주제어를 설정하게 되며 핵심 어휘를 중심으로 내용의 흐름을 상세화하게 된다. 따라서 핵심 어휘는 언어적 특성을 드러내는 중핵적인 역할을 담당한다고 볼 수 있다.

또한 내러티브의 전개 방식에 있어서 특징적인 속성은 ‘사건’으로, 필

---

1) 인지심리학 연구에 따르면 하나의 어휘가 지닌 개념(의미)은 사고의 기본 단위로 작동한다.

자는 연상된 핵심 어휘, 연계 어휘, 문법 요소를 사용하여 사건에 대해 묘사하고 내용을 전개할 것이다. 특히 사건의 나열, 혹은 사건과 사건의 관계를 나타내는 문법 요소 중에 시제와 상은 텍스트 전체의 응결성에 큰 영향을 줄 것이며, 사건을 묘사하는 데 중요한 요소로 작용한다. 요컨대 핵심 어휘와 연계 어휘, 그리고 이들을 구성하는 문법 요소는 내러티브 텍스트의 장르성을 드러내는 언어적 장치로 기능할 수 있다. 따라서 텍스트에서 어휘와 문법 요소를 어떻게 사용했는지를 살펴보는 것은 외국인 학습자가 텍스트의 맥락을 적절하게 고려했는지를 판단하는 지표가 될 수 있을 것이다. 이러한 논의를 통해 내러티브와 관련한 교수·학습 내용의 근거를 마련하는 데 보탬이 될 것으로 기대한다.

## 2. 선행 연구

한국어 교육 분야에서 내러티브 텍스트에 대한 연구는 대부분 구어 내러티브에 한정되어 이루어졌으며 작문에 대해서는 미비하게 연구가 이루어져 왔다. 이마저도 내용 또는 구조 분석, 오류 분석을 중심으로 이루어져 텍스트의 장르성과 그 언어적 특성에 주목한 연구는 소수에 불과하다.

지금까지 내러티브 텍스트의 구조와 언어적 특성에 대한 연구는 강은실(2007), 오지혜(2008, 2016), 신필여(2010), 심은지(2011), 조국현(2013), 조은(2015), 한송화(2014, 2015), 유민애(2015, 2017) 등이 있다. 이 중 한국어 학습자가 생산한 텍스트를 바탕으로 내러티브의 장르성과 언어적 표현을 관련지어 연구한 조은(2015), 한송화(2015), 유민애(2015, 2017)의 연구가 대표적이다. 특히 조은(2015)은 외국인 한국어 학습자의 내러티브 텍스트의 구조에 따른 시제와 상의 사용 양상을 분석하였는데, 이는 내러티브에서 나타나는 언어적 표현의 특성을 밝히려고 했다는 점에서 의미가 있다. 다만 학습자들이 범하는 오류의 유형에 주목하여 이루어졌으며, 외국인 모어 화자의 텍스트 등과 같은 비교 대상 없이 숙달도가 일정하



지 않은 한국어 학습자의 텍스트만을 대상으로 분석이 이루어졌다는 한계점이 있다. 또한 조은(2015)과 같이 특정 텍스트에서 나타나는 문법 요소의 특성을 밝히고자 하는 연구들은 언어적 표현을 추출하는 과정이나 분석하는 과정에서 주로 정성적 분석 방법이 사용되었다. 정성적 분석은 분석자의 전문성과 훈련이 기반이 되어야 하기 때문에 객관적인 평정이 쉽지 않다. 물론 텍스트에서 사용된 언어적 특성을 살핀다는 것은 학습자의 인식을 탐구하는 과정이기에 정성적 평가는 쓰기 평가에서 필수적이지만 분석 대상의 수가 많거나 대단위의 텍스트를 분석하기에는 어려움이 따른다. 이러한 한계를 극복하고자 텍스트에 나타난 언어적 표현들을 통계적으로 분석하는 연구들은 여전히 진행 중이며, 이와 관련하여 한송화(2015), 유민애(2015, 2017) 등은 코퍼스를 구축하고 텍스트의 구조와 언어 사용 양상을 분석했다는 점에서 의미를 찾을 수 있다.

실제 텍스트에 나타난 언어적 특성을 객관적이고 통계적으로 분석한다는 것은 쉽지 않다. 텍스트에 나타난 언어를 살펴본다는 것은 필자의 의도를 파악하는 일이며, 이는 섬세한 해석을 요하는 작업이기에 통계적 분석 기법을 사용하는 경우 놓치는 부분이 생겨 본래의 목적과는 다르게 계량적 결과만이 도출될 수 있기 때문이다. 이러한 이유로 텍스트 분석 기술과 방법론에 대한 연구는 보완을 거듭하면서 꾸준히 성장하고 있다. 국어 교육 분야에서 쓰기 자동화 채점이나 한국어 교육 분야의 연구 동향을 분석하는 연구 등에서 빅데이터를 활용하는 시도가 활발하게 나타나며 특정 유형의 텍스트와 문법 요소들의 특성을 알아보기 위해 최신 분석 기술을 접목하는 시도도 늘고 있다.<sup>2)</sup>

그간 영어 교육에 비해 한국어 교육 분야에서 분석 기술의 도입이 어려웠던 이유 중 하나는 한국어가 가진 교착어적 특성과 어순, 띄어쓰기<sup>3)</sup>

---

2) 최근 텍스트 마이닝 기법을 활용하여 학습자의 텍스트를 살핀 연구(장연지 외, 2020; 채윤미·길호현, 2020; 김은정 외, 2021)가 조금씩 이루어지고 있는데, 새로운 텍스트 분석 기술을 도입하여 유의미한 결과를 도출했다는 점에서 다양한 텍스트 관련 연구에 시도될 수 있는 토대를 제공하였다.

등의 특수성 때문일 것이다. 이러한 이유로 EFL에서 사용하던 기준을 그대로 가져오거나 연구자가 임의의 기준을 가지고 전산화 작업을 하는 경우가 많았다. 이러한 점에서 본 연구는 최신 정보 처리 기술이 한국어 교육 분야의 연구에 활용될 수 있다는 가능성을 타진하기 위해 그 방법을 모색하고 시도해 보고자 한다.

### 3. 연구 방법

#### 3.1. 연구 절차 및 내용

본 연구는 외국인 한국어 학습자와 한국인 모어 화자의 내러티브 텍스트에 나타난 핵심 어휘와 문법 요소의 사용을 비교 분석하기 위해 분석 대상인 텍스트 수집과 선정, 텍스트 구조 분석의 과정을 거칠 것이다. 그리고 텍스트 마이닝의 언어 네트워크 기법을 활용하여 구조에 따른 언어적 특성을 파악하여 한국인과 외국인의 차이를 파악해 보고자 한다.

이를 위해 한국인 46명과 외국인 43명을 대상으로 내러티브 텍스트를 수집하였다. 연구에 참여한 한국인 모어 화자는 A대학교에 재학 중인 2학년과 3학년 학생들이며 외국인 학습자의 경우 같은 학교 대학원에 재학 중인 유학생들로 한국어능력시험(TOPIK)에서 5급 또는 6급을 받은 바 있다<sup>4)</sup>. 참여자들에게 연구 내용을 설명한 후 ‘인생에서 가장 후회하는 일’이라는 주제<sup>5)</sup>로 같은 환경에서 1000자 내외의 텍스트를 생산하도록

---

3) 가령 문장 토큰화(Sentence Tokenization)의 경우, 띄어쓰기 단위가 영어처럼 독립적인 단어라면 그대로 토큰화를 하면 되겠지만, 한국어는 어절이 독립적인 단어로 구성되는 것이 아니라 조사 등이 붙어있는 경우가 많아서 이를 모두 분리하는 작업이 필요하다.

4) 외국인 학습자의 국적은 중국, 베트남, 라오스이며 한국 거주 기간은 1년에서 3년 사이다.

5) 주제를 선정할 때 ‘후회’는 누구에게나 있을 법한 보편적인 주제이면서도 사건을 중심으로 경험을 기술할 수 있다는 점, 또한 한국의 대학 부설 한국어 교

하였다. 수집 후 검토를 통해 분량이 너무 적거나 많은 경우, 주제와 관련 없는 경우, 어휘와 문법의 오류가 많이 나타난 텍스트는 선정에서 제외하였으며 최종적으로 한국인의 텍스트 40편, 외국인의 텍스트 40편을 분석 대상으로 선정하였다.

핵심 어휘와 문법 요소를 분석하기에 앞서 한국인 모어 화자와 외국인 학습자의 텍스트에 나타난 구조 출현의 차이를 살펴보고자 80편의 텍스트를 대상으로 구조 분석을 하였다. 구조 분석의 틀을 정하기 위해 선행 연구들을 검토한 결과 미시적인 구조보다는 거시적인 구조가 분석의 틀로 적합하다고 판단하였다. Labov(1972)에서 제시한 ‘개요(Abstract) - 배경 소개(Orientation) - 사건(Complication action) - 평가(Evaluation) - 결과(Resolution) - 종결(Coda)’이나 강은실(2007)의 ‘개요 - 배경 - 사건 - 평가 - 교훈’ 등과 같이 미시적으로 구조를 나눈 경우<sup>6)</sup>는 연구의 목적이 내러티브 텍스트의 구조 분석이거나 이야기하기의 원형이 구조적으로 실현되는 방식을 살펴보는 것이라면 필요하겠지만, 본 연구는 구조에 따른 언어적 특성을 분석하는 것이 목적이므로 거시적 구조가 분석의 틀로 적절하다고 보았다. 또한 텍스트 마이닝의 언어 네트워크 분석을 하기 위해서는 각 구조마다 일정량 이상의 단어 수가 필요하다<sup>7)</sup>. 이러한 이유로 본 연구는 한송화(2015)의 ‘도입부 - 사건전개부 - 결말부’를 내러티브 구조의 분석 틀로 설정하였으며 구조별로 포함되는 내용은 다음과 같다.

---

육 기관에서 출판한 교재 대부분에서 해당 주제가 쓰기 과제로 제시되어 있기 때문에 외국인 학습자들이 쓰기 경험을 했을 거라는 점 등을 고려하였다. 또한 내러티브 텍스트에서 중심이 되는 것이 사건과 시간성이라는 점에서 볼 때 과거에 일어난 특정 사건을 서술하는 과정에서 시제 등 문법 요소의 실현 양상을 분석하기에 적합하다고 판단하여 주제로 선정하였다.

- 6) 이 외에도 조은(2015), 유민애(2017) 등이 있다.
- 7) 미시적 구조에 따라 언어 네트워크 분석을 할 경우 특정 구조부를 생략하거나 한 구조부에 한두 문장만으로 짧게 기술한 사례들이 많이 있어 구조별 언어적 특성을 살펴보기에 어렵다고 판단하였다.

<표 1> 한송화(2015: 516)의 내러티브 구조

구조	내용
도입부	이야기를 시작하기에 앞서 사건 주제를 진술
사건 전개부	<ul style="list-style-type: none"> <li>• 사건이 일어나게 된 배경, 사건이 일어난 장소나 시기, 등장인물들에 대한 상황적 배경을 설명</li> <li>• 주요 사건이나 등장인물들의 행동, 각 인물들의 사건에 대한 감정을 기술</li> <li>• 갈등과 문제가 해결되고 매듭지어짐</li> </ul>
결말부	사건에 대한 필자의 해석과 평가, 이로 인한 필자의 삶의 변화 등을 기술

Beaugrande(1981)을 비롯한 많은 연구에서 텍스트의 응결성은 시제·상과 관련이 있음을 주장한 바와 같이 내러티브 텍스트는 경험을 기술하는 과정에서 시간에 관한 표현을 써야 완전한 이야기를 재현할 수 있기 때문에 시제·상과 같은 문법 요소가 많이 사용된다. 본 연구에서도 이 연구들과 맥을 같이 하며 시제·상 분석의 틀 마련을 위해 선행 연구들을 검토하였다. 그리고 한송화(2015), 조은(2015)을 참고하여 <표 2>, <표 3>과 같은 분석 틀을 설정하였다.

<표 2> 내러티브 텍스트의 시제 분석을 위한 틀 (조은, 2015: 38)

	과거 시제	현재 시제	미래 시제
종결어미	-었-, -았었-, -았겠-	-는/니-, -는-, ∅	-겠-
연결어미		∅	
전성어미	-(으)ㄴ[동사], -던[형용사/-이다]	-(으)ㄴ[형용사/-이다] -으(ㄴ), ∅	-(으)ㄴ
기타			-(으)리- -(으)ㄴ 것이다

<표 3> 내러티브 텍스트의 상 분석을 위한 틀

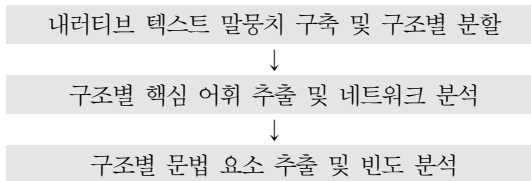
	보조동사	연결어미
완료상	-고 나다, -고 말다, -고 있다 <sub>2</sub> (계시다 <sub>2</sub> ), -어 내다, -어 놓다, -어 두다, -어 버리다, -어 있다(계시다), -어 지다, -어 치우다	-다가, -어다가, -자마자
진행상	-곤 하다, -어 가다/오다, -고 있다 <sub>1</sub> (계시다 <sub>1</sub> ), -어 대다, -는 중이다, -는 길이다	-면서
예정상	-게 되다	-고자, -도록, -(으)러, -(으)려고

내러티브 텍스트의 구조 분석은 경력 10년 이상의 한국어 교사 2명과 함께 진행하였으며 시제·상 분석을 위한 틀을 마련하는 과정에서도 이들에게 의견을 구하였다.

### 3.2. 분석 도구 및 방법

분석은 두 단계로 진행하였으며 전체적인 진행 절차는 다음과 같다.

<표 4> 분석 절차



#### 3.2.1. 말뭉치 구축 및 구조별 분할

한국인 모어 화자 40명과 외국인 학습자 40명이 작성한 텍스트를 각각 별도의 말뭉치로 구축하였다. 개별 텍스트에 내러티브 구조를 표시하는 표지([도입], [사건 전개], [결말])를 삽입하여 구조에 따른 분할이 용이하도록 하였다. 또한 구조에 따른 차이를 비교할 수 있도록 구조별로 분할된 하위 말뭉치를 각각 별도로 구축하였다. 말뭉치 구축 후 글자 수, 문단 수, 문장 수, 어절 수, 구조별 어절 수 등의 기초 자료를 산출하였다.

말뭉치 구축 후에는 형태소 분석을 실시하였다. 형태소 분석은 분석의 신뢰도가 UTagger를 사용하였으며 체언에 ‘-하’, ‘-되’가 붙어서 형성되는 용언의 경우 이를 분리하지 않고 하나의 단어로 처리하도록 하였다.

#### 3.2.2. 구조별 핵심 어휘 추출 및 네트워크 분석

구조에 따른 내용 구성 양상의 차이를 비교하기 위해서 실질 형태소 중 의미 기능을 하는 명사(NNG), 대명사(NNP), 수사(NP), 동사(VV), 형용사(VA), 일반 부사(MAG), 접속 부사(MAJ)를 추출하였다. 이와 같은 과정을 통

해 불용어 제거 등 별도의 전처리 과정 없이 내용 어휘를 선정할 수 있다.

내용 어휘 선정 후 문서와 어휘의 행렬(DTM: Document Term Matrix)을 생성하였다. 단어 사용 양상의 차이를 보다 분명히 드러내기 위해 DTM 생성 과정에서 Tf-idf 값을 가중치로 부여하였다. 그런데 이렇게 생성된 DTM은 빈칸이 매우 많은 희소 행렬이므로 일정 수준 이상이 빈칸이 되는 단어들을 일괄 삭제하는 방식으로 핵심 어휘의 수를 제한하였다. 도입과 결말 부분은 분량이 적기 때문에 90% 이상이 빈칸인 단어들을 삭제하였으며 사건전개 부분에서는 80% 이상이 빈칸인 단어들을 삭제하였다.<sup>8)</sup>

구조별로 DTM을 생성하고 핵심 어휘를 추출한 후 각 단어 간의 상관계수로 연결 관계를 산출하였다. 연결 선의 수를 통제하기 위해 도입과 결말 부분은 상관 계수가 0.2 미만인 관계를 삭제하였으며, 사건 전개 부분은 0.25 미만인 관계를 삭제하였다. 단어 및 관계를 삭제하기 위한 수치는 한국인 집단과 외국인 집단 말뭉치에 모두 동일하게 적용하였다.<sup>9)</sup>

핵심 어휘의 선정과 단어 간 상관 계수가 이루어진 후 구조별 핵심 어휘

8) DTM은 그 특성상 희소행렬이 생성될 가능성이 크기 때문에 적절한 방법으로 노드의 수를 조절할 필요가 있다. 일반적으로 일정 희소성(sparsity) 이상인 단어를 지우거나 특정 빈도 이상인 단어로 한정하는 방식을 사용하는데, 이 연구에서는 희소성을 기준으로 적용하였다. 예를 들어, 도입 부분은 37개 문서에 총 520개 단어가 사용되었고 전체 희소성은 95%로 나타났다. 이중 희소성이 90% 이상인 단어를 삭제하면 총 44개의 단어가 남고 희소율은 76%로 감소한다. 마찬가지로, 결말 부분은 총 40개 문서에 670개의 단어가 사용되었는데, 희소성이 90% 이상인 단어를 삭제한 결과 65개의 단어가 남았다. 한편 사건전개 부분은 총 40개의 문서에 1711개의 단어가 사용되었는데, 이 경우 희소율이 90% 이상인 단어를 삭제하면 196개의 단어가 남는다. 196개의 단어 역시 네트워크 분석을 하기에는 지나치게 노드 수가 많으므로 삭제하는 희소성 기준을 80%로 강화하여 총 73개의 단어를 산출하였다.

9) 이 연구에서 노드를 연결하는 엣지는 각 단어의 피어슨 상관계수로 설정하였다. 그런데 희소성이 높은 단어를 삭제하고 남은 단어들은 서로 긴밀하게 연결되기 때문에 엣지의 수가 지나치게 많아진다. 예를 들어, 도입부의 경우 44개의 단어에서 총 946개의 엣지가 생성되는데 낮은 상관관계의 기준인 0.2 미만을 0으로 처리했을 때 엣지 수가 148개로 조정된다. 노드와 엣지의 수를 결정하는 객관적 기준이 별도로 있지 않기 때문에 네트워크 분석이 가능한 수준에서 설정하는 한계가 존재한다.

네트워크 구조를 시각화하였다. 각 노드의 크기는 연결 중심성(degree centrality)으로 나타냈으며 노드의 색은 하위 그룹에 따라 다르게 형성되도록 하였다. 하위 그룹은 R의 tidygraph 패키지에서 제공하는 group\_infomap 함수를 통해 산출하였다.

### 3.2.3. 구조별 문법 요소 추출 및 빈도 분석

구조별로 사용되는 문법 요소의 양상 차이를 분석하기 위해 형태소 분석이 된 말뭉치에서 분석하고자 하는 문법 요소를 추출하였다. 문법 요소는 앞에서 <표 2>, <표 3>으로 제시한 항목을 사용하였다. 다만 검색의 대상이 형태소 분석이 된 말뭉치이기 때문에 문법 항목도 형태소가 분석한 형태로 바꾸어 줄 필요가 있다. 즉 과거시제 연결어미 ‘-었겠-’의 경우 ‘었/EP+ 겠/EP’와 같이 수정하여 검색을 해야 한다. 이렇게 형태소 분석이 된 상태로 문법 요소를 검색하면 어절 단위 이상의 문법 요소를 검색할 수 있다는 장점이 있다. 예를 들어, 완료상을 나타내는 보조 동사 ‘-고 말다’와 같은 형태는 기존의 단어 단위 텍스트 마이닝 분석에서는 삭제되는 것이 일반적이었으나 ‘고/EC 나\_01/VX’와 같은 형태로 검색하여 빈도를 추출할 수 있다. 이처럼 분석하고자 하는 문법 요소들을 모두 형태소 분석하여 리스트로 만들고, 분석 대상 말뭉치에서 이 리스트의 문법 요소에 대한 발생 빈도를 각각 산출하였다. 분석의 정확성을 기하기 위해 형태소 분석 말뭉치를 대상으로 개별 문법 요소에 대한 빈도를 연구진이 전수 조사하여 확인하였다.

## 4. 내러티브 텍스트의 구조별 분석 결과

한국인 모어 화자와 외국인 학습자가 생산한 내러티브 텍스트의 구조를 본격적으로 살펴보기에 앞서 각 텍스트의 문장 및 문단 수, 총 어절 수 및 각 구조 부분의 어절 수를 분석하였으며 그 결과는 다음과 같다.

&lt;표 5&gt; 내러티브 텍스트의 나타난 글자 수와 구조별 어절 수

	글자 수	순수 글자 수	문단 수	문장 수	어절 수	도입부 어절 수	사건전개 어절 수	결말부 어절 수
한국인	932.4	681.4	3.8	20	235.4	36.9	149.3	49.4
외국인	951.5	688.6	4.1	20.7	245.4	35.1	163.8	46.5

한국인과 외국인의 텍스트에 나타난 글자 수, 문장 수, 어절 수를 비교한 결과 외국인의 텍스트에서 약간 많은 것으로 나타났는데 큰 차이는 아니었다. 텍스트에 나타난 평균 문단 수 역시 3.8개와 4.1개로 큰 차이를 보이지 않았다. 한국인 텍스트의 경우 사건전개부가 한 문단 또는 두 문단으로 이루어져 있었으며 외국인 집단 역시 비슷한 양상을 보였으나 세 문단으로 구성된 경우도 가끔 나타났다. 두 집단 모두 도입부와 결말부의 문단은 각각 한 문단으로 이루어진 경우가 대부분이었다. 주목할만한 점은, 전체 어절 수는 외국인이 10어절 정도 많은 것으로 나타났는데 사건 전개 부분에서만 외국인 텍스트의 어절 수가 많고 도입과 결말 부분의 어절 수는 한국인 텍스트에서 더 많이 나타났다는 점이다. 이는 외국인이 한국인에 비해 도입과 결말보다 사건 전개에 더 많은 내용을 서술했다는 것을 보여준다.

구조를 분석한 결과 80편의 텍스트 중 73편의 텍스트에서 모두 ‘도입부-사건전개부-결말부’의 구조를 보였으며 한국인 5편, 외국인 2편의 텍스트에서는 ‘도입부’가 생략된 ‘사건전개부-결말부’의 구조를 보였다. 도입부가 나타나지 않은 텍스트는 아래와 같이 텍스트의 첫 문장부터 배경을 소개하거나 주요 사건을 서술하는 ‘사건전개부-결말부’의 구조로 나타났다.

- ① 나는 대학교 때 중국에서 한국어를 배웠다. 우리 한국어 학과의 모든 학생이 3학년 때 한국에서 교환을 해야 한다. (F35)
- ② 2019년 10월 중순 무렵, 독서실에서 공부를 하고 있던 나는 아버지께 카톡을 보냈다. ‘나 공부하기 싫고, 너무 힘들어.’ (K17)



①은 도입 없이 사건전개부의 한 부분인 배경을 제시하며 첫 문장을 서술한 예이며 ②는 도입과 배경 설명 등을 생략한 채 사건에 대해 서술하고 있다. 이는 독자의 흥미를 끌기 위해 사건을 먼저 제시한 전략으로 보이며 이러한 양상은 외국인보다 한국인의 텍스트에서 더 많이 나타났으나 유의미하다고 말할 수 있을 정도의 큰 차이는 아니었다.

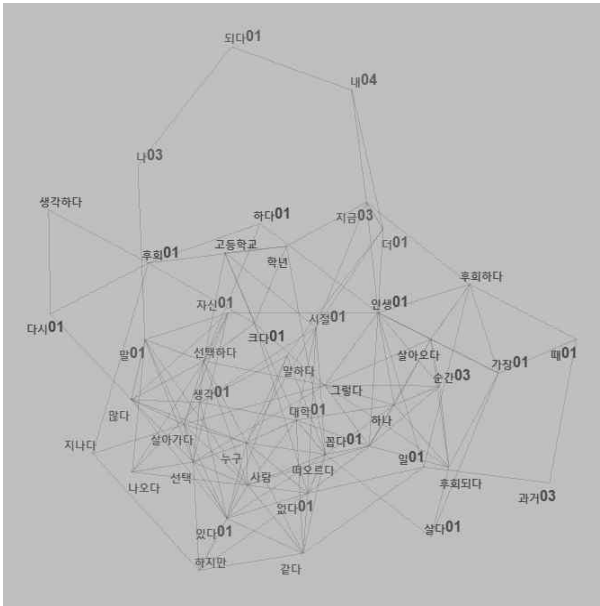
다음으로 각 구조에 나타난 핵심 어휘와 시제·상의 사용 양상을 분석하였다.

#### 4.1. 핵심 어휘 분석

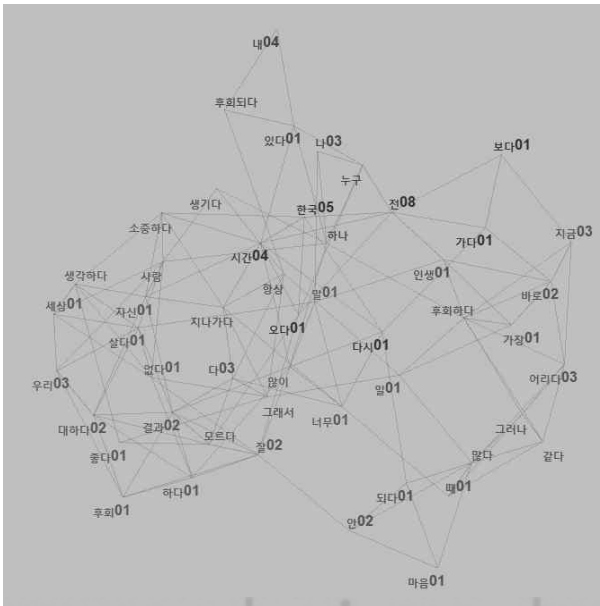
텍스트에 나타난 핵심 어휘를 분석하기 위해 먼저 두 집단의 텍스트에 나타난 어휘의 빈도를 산출하였다. 전체 및 구조별로 가장 많이 등장한 어휘를 추출한 후, 이 어휘 목록을 바탕으로 구조별 핵심 어휘의 네트워크 구조를 시각화하여 나타내 보았다. 언어 네트워크에서는 핵심 어휘들이 그 관계에 따라 선으로 연결되어 여러 의미 군집 형태로 나타난다. 이 중에서 같은 군집의 어휘들은 더 긴밀한 의미적 관계를 갖는데, 이렇게 가시화된 데이터를 통해 직관적으로 주제와 필자의 표현 의도 및 수사적인 전략을 파악할 수 있다. 핵심 어휘 네트워크를 활용해 내러티브 텍스트의 각 구조에 나타난 두 집단의 특징을 살펴보았다.

##### 4.1.1. 도입부

도입부는 내러티브 전체의 내용을 1~3문장으로 요약하여 제시하거나 본격적인 사건에 대해 설명하기 전에 주제를 언급함으로써 독자의 흥미를 일으키도록 하는 기능을 한다. 한국인과 외국인 집단의 내러티브 텍스트 도입 부분의 어휘 네트워크를 분석한 결과 다음과 같은 결과를 얻을 수 있었다.



[그림 1] 한국인 모어 화자의 텍스트 도입부에서 추출한 핵심 어휘 네트워크



[그림 2] 외국인 학습자의 텍스트 도입부에서 추출한 핵심 어휘 네트워크

먼저 한국인 집단의 도입부에 나타난 핵심 어휘 네트워크에서는 두 가지 연결선이 특징적이다. 첫 번째는 [그림 1]의 왼쪽 하단에 나타난 어휘들 중 ‘사람, 누구, 살아가다, 선택, 있다’의 연결이 보이는 것을 확인할 수 있다. 이는 텍스트에서 필자가 ‘사람은 누구나/누구든지 인생에서 후회되는 일이 있다’와 같은 구문을 많이 사용했기 때문인 것으로 볼 수 있다. 두 번째는 [그림 1]의 오른쪽 중간부터 하단까지 분포한 어휘들 중 ‘인생, 살아오다, 살다, 순간, 후회하다, 후회되다’의 연결을 확인할 수 있는데 이 역시 ‘인생을 산다, 매 순간 선택을 한다, 후회를 한다’와 같은 구문을 사용한 것으로 추측할 수 있다. 이는 글의 시작 부분에서 주제인 ‘후회’의 보편성을 강조하고, 사건에 대한 설명을 시작하기 전 텍스트 전체를 아우를 만한 개요의 성격을 가진 문장을 기술하기 위한 의도로 파악된다. 실제 한국인 집단의 도입 부분에서 아래와 같은 구문이 많이 등장하는데 이는 핵심 어휘의 네트워크에 나타난 양상과 일치한다.

- ③ 사람은 누구나 살아가면서 후회되는 일이 있을 것이다. (K03)
- ④ 인생은 선택이라는 말이 있다. (중략) 그렇기에 많은 사람들은 자신의 선택에 후회하며 살아간다. (K04)
- ⑤ 인생은 선택의 연속이다. 사람들은 인생을 살아가면서 매 순간마다 선택을 하게 된다. (중략) 사람들은 그 선택에 대한 후회를 하기도 한다. (K24)

③~⑤와 같이 한국인 모어 화자의 텍스트에서는 후회하는 일이 누구에게나 있다는 문장이 도입부에 자주 출현한다. 특히 ④의 ‘~이라는 말이 있다’와 같은 인용문을 통해 인생이 선택의 연속이며 그 과정에서 후회는 불가피하다는 구문이 많이 등장하는데 이는 [그림 1]의 네트워크에서 오른쪽 하단의 핵심 어휘 ‘말’과 연결되는 것으로도 확인할 수 있다. 이러한 도입은 누구나 필연적으로 느꼈을 후회의 감정을 언급함으로써 독자의 공감을 불러일으킬 수 있을 뿐만 아니라 광고나 보편적으로 알고 있는 문구로 독자의 주의를 끄는 효과를 얻을 수 있다.

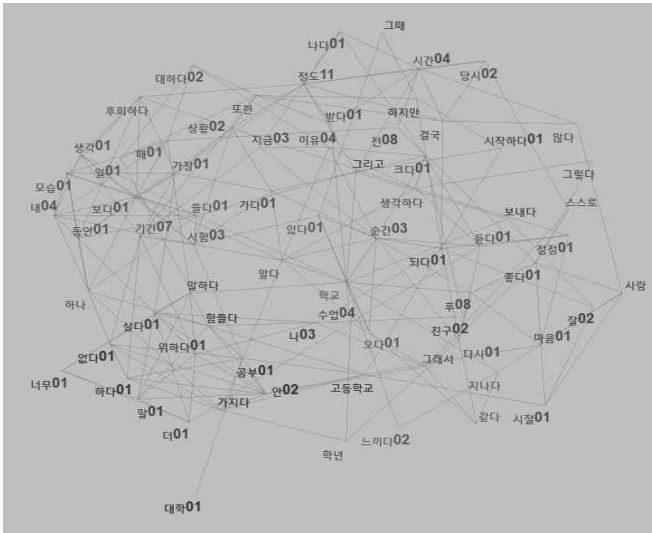
외국인 집단의 도입부에 나타난 핵심 어휘 네트워크의 특징으로는 [그림 2]에 나타난 왼쪽 중간부터 아래쪽으로 연결된 ‘생각하다’, ‘좋다’, ‘모르다’ 등의 어휘가 ‘세상, 우리, 후회’와 연결되는 것을 확인할 수 있는데, 이는 ‘좋다’라는 평가 형용사, ‘모르다’와 ‘생각하다’라는 인지 동사, ‘우리’라는 1인칭 대명사로 필자가 자신의 감정, 판단, 즉 주관적인 정서를 드러내었음을 추측할 수 있다. 이와 같은 핵심 어휘가 나타난 외국인 학습자의 실제 도입부를 살펴보면 다음과 같다.

- ⑥ 내가 한참 생각했는데 후회하는 일이 떠오르지 않았다. 왜냐면 지금까지 내가 한 모든 일이 다 내가 스스로 선택한 일이었기 때문이다. (F29)
- ⑦ 나에게는 후회하지 않는 일이 좋은 편이고, 생각해 보면 후회하는 일이 좋은 편이라고 생각한다. (F30)

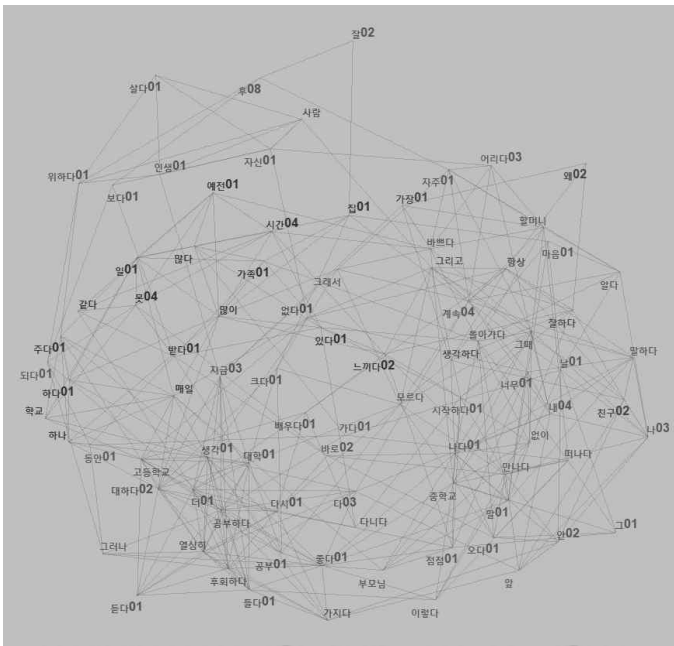
위 ⑥, ⑦과 같이 외국인 학습자들의 텍스트 도입부에서는 1인칭 대명사를 사용에 자신의 감정이나 주관적인 견해를 표현하는 구문을 다수 확인할 수 있다. [그림 2]에서 ‘생각하다’, ‘살다’ 등의 어휘가 왼쪽 하단에 나타난 ‘우리’뿐만 아니라 중간 상단의 ‘나’와 ‘내’ 같은 1인칭을 나타내는 명사로 연결되는 것을 볼 때 외국인 학습자들은 ‘우리~ 살다’, ‘내가 생각했는데’, ‘나에게는 ~라고 생각한다’와 같은 표현을 자주 사용하여 서술한 것으로 보인다. 이는 한국인 집단의 네트워크에는 나타나지 않은 특징으로 외국인 집단의 경우 자신의 경험을 서술한다는 내러티브 텍스트의 장르적 특성을 인식하여 1인칭 대명사를 과도하게 사용한 것으로 해석할 수 있다.

#### 4.1.2. 사건전개부

사건전개부는 사건이 이루어지기 이전의 당시 상황이나 장소, 인물 등을 설명하는 배경 소개와 사건을 서술하는 구조부로 내러티브 텍스트의 중심에 해당하는 부분이다. 한국인과 외국인 집단의 내러티브 텍스트 사건전개부의 핵심 어휘 네트워크는 [그림 3], [그림 4]와 같이 나타났다.



[그림 3] 한국인 모어 화자의 텍스트 사건전개부에서 추출한 핵심 어휘 네트워크



[그림 4] 외국인 학습자의 텍스트 사건 전개부에서 추출한 핵심 어휘 네트워크

[그림 3]을 보면 한국인 집단의 경우, 왼쪽 상단의 어휘 연결 분포와 왼쪽 하단의 어휘 연결 분포가 특징적인데 ‘시험, 기간, 가장, 후회하다’의 연결과 ‘고등학교, 대학교, 힘들다, 공부, 안, 하다’ 연결을 확인할 수 있다. 사건전개부의 핵심 어휘 네트워크에 나타난 어휘 연결을 통해 사건의 주제와 흐름을 파악해 보자면 한국인 집단에서는 시험 기간에 후회한 일, 학교에서 공부를 안 한 일 등을 소재로 사건을 서술한 것으로 파악된다. 그리고 실제 한국인의 텍스트를 확인해 본 결과 다음과 같은 문장을 발견할 수 있었다.

- ⑧ 고등학교 시절에는 학습 계획을 세우지 않아 시험 기간에는 혼란이 왔으며, 자주 놀고 싶은 유혹에 빠져 공부를 던져놓기도 했다. (K2)
- ⑨ 그럼에도 불구하고 후회하는 일을 떠올려보면 고등학교 2학년 시절이 제일 후회가 된다. 남들이 모두 목표를 가지고 내신 공부를 할 때 나는 ‘목표의식이 없는데 공부를 해서 뭐 하나’라는 말도 안 되는 생각에 사로잡혀 공부를 안했다. (K11)

외국인 집단의 경우 [그림 4]에서 왼쪽 하단에서 ‘고등학교, 대학, 더, 열심히, 공부하다, 후회하다’의 연결을 확인할 수 있다. 또한 오른쪽 상단에서 ‘어리다, 할머니, 돌아가다’의 연결과 중앙 10시 방향 근처에 ‘가족, 예전, 시간, 못’의 연결을 볼 수 있다. 외국인 집단의 경우도 한국인 집단과 비슷하게 학교에서 더 열심히 공부하지 않은 일과 관련하여 사건을 서술한 것으로 나타났으나 그보다는 가족의 임종을 지키지 못했거나 가족과 시간을 함께 보내지 못해 아쉬웠던 경험이 주된 사건의 내용으로 서술된 것으로 나타났다는 점에서 한국인 집단과 차이를 보였다.

또한 한국인 집단에서는 외국인 집단에 비해 가족 관련 어휘가 적게 등장하였는데, 대신 ‘생각’, ‘시간’, ‘학년’, ‘과제’ 등의 어휘가 나타났다. 이와 관련하여 실제 텍스트를 살펴본 결과 외국인 집단에서 인간관계를 주제로 사건을 서술한 경우가 한국인 집단에 비해 2배 가까이 많은 것으

로 확인되었다. 반면, 한국인 집단에서는 공부 또는 학업과 관련된 사건의 원인을 성격이나 성적에서 찾으며 후회하는 내용이 많은 것을 알 수 있었다. 다음은 집단별 사건전개부의 주제를 살펴본 결과이다.

<표 6> 사건전개부의 주제

		인간관계	성격	학업	취미 관련	기타
한국인	10	가족: 4 친구: 6	14	12	3	1
외국인	19	가족: 14 친구: 3 타인: 2	5	11	2	3

실제 텍스트를 살펴본 결과 가족과 관련된 사건을 서술한 경우가 많이 발견되었으며 특히 할머니와 관련된 일화가 많이 등장함을 알 수 있었다.

- ⑩ 지금 인생에서 가장 후회하는 일은 할머니께서 돌아가시기 전에 옆에 자주 있지 못한 것이다. (F04)
- ⑪ 어릴 시절에 할머니께서 시각 잃어버린 전에 할머니와 같이 살았다. 고생 많이 하신 할머니께서 돌아가신 그날은 내 인생에 가장 힘들고 고통스러운 날이었다. (F24)
- ⑫ 예전에는 코로나 때문에 집에 있는 동안에 사소한 일로 부모님과 언쟁을 하기도 하고 빨리 한국에 가서 공부도 하고, 빨리 부모님 옆에 벗어나야 한다는 생각까지 들었다. 그러나 시간이 갈수록 가족과 함께 보내는 시간이 얼마나 소중한지 깨닫지 못하게 되었다. (F36)

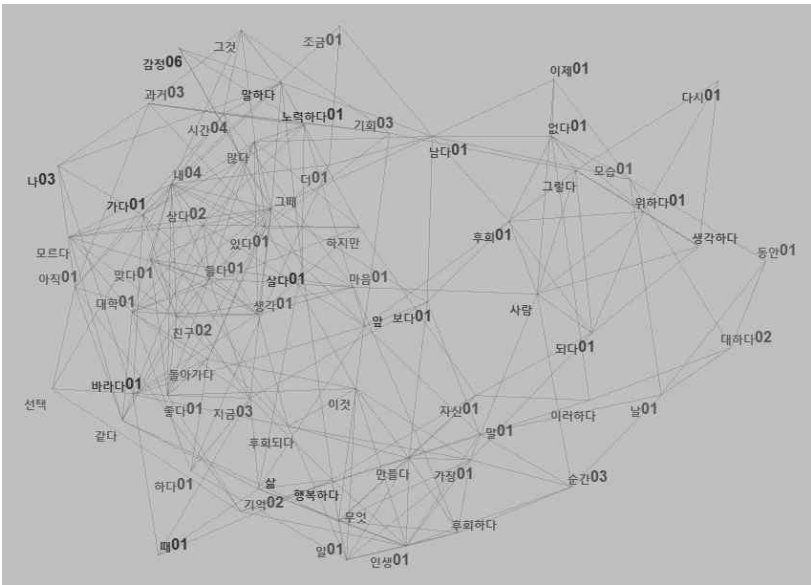
외국인 집단의 텍스트에서 많이 나타난 조부모의 임종과 추억에 관한 내용은 학습자의 성장 배경에서 비롯된 것으로 짐작된다. 연구에 참여한 유학생 상당수가 중국인이라는 점에서 봤을 때, 어린 시절 맞벌이 또는 도시로 일을 하러 간 부모 대신 조부모의 손에서 자란 이들의 성장 배경

이 과거의 일을 회상하는 과정에 반영되었을 것이다. 또한 현재 유학생이라는 신분적 상황과 팬데믹으로 인해 가족을 만난 지 오래되고 그리워하는 마음이 사건의 주제 선정에 반영되었으며 이러한 결과가 언어적 특성으로 나타난 것으로 보인다.

사건 전개부는 어휘 네트워크를 통한 한국인 집단과 외국인 집단의 주제 및 글의 흐름의 차이가 가장 잘 나타난 부분으로 삶의 경험을 이야기하는 내러티브의 특성이 잘 나타났다고 볼 수 있다.

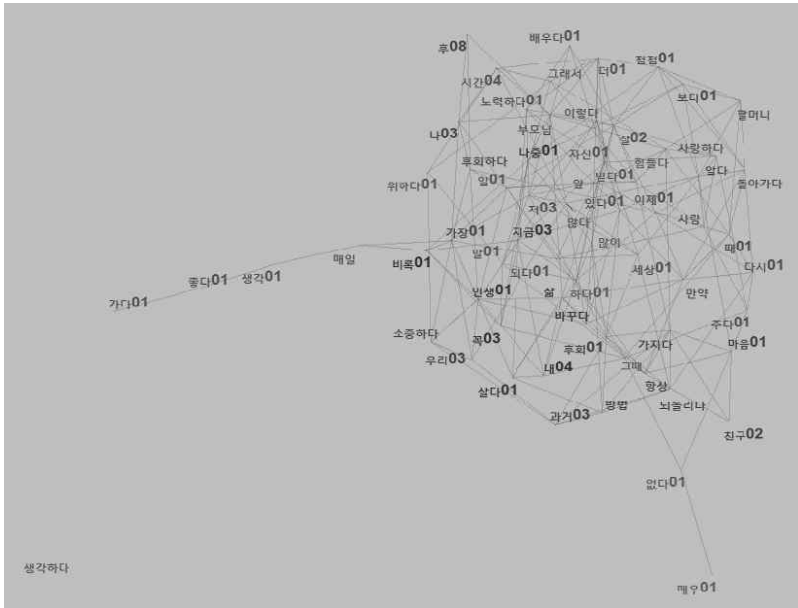
#### 4.1.3. 결말부

결말부는 사건에 대한 필자의 감상이나 평가, 그리고 사건의 교훈이나 사건과 현재에 끼친 영향 등이 나타나며 텍스트를 마무리 짓는 부분이다. 한국인과 외국인 집단의 텍스트 결말부의 핵심 어휘 네트워크는 [그림 5], [그림 6]과 같이 나타났다.



[그림 5] 한국인 모어 화자의 텍스트 결말부에서 추출한 핵심 어휘 네트워크





[그림 6] 외국인 학습자의 텍스트 결말부에서 추출한 핵심 어휘 네트워크

한국인 집단의 텍스트 결말부에 나타난 핵심 어휘 네트워크를 살펴본 결과, 먼저 [그림 5]에서 왼쪽 중간에서 아래로 이어지는 ‘내, 그때, 돌아 가다’를 볼 수 있는데 이는 한국인 필자가 마지막 부분에서 다시 과거로 돌아간다는 가설을 통해 자신의 후회 감정을 표출한다는 내용을 담아 서술했음을 추측할 수 있다. 이에 대한 구체적 양상을 확인하기 위해 한국인 집단의 결말 부분을 살펴보았다.

- ⑬ 공상적인 이야기지만, 만약 과거로 돌아갈 수 있다면 나는 2017년의 휴학을 바꾸고 싶다. (K01)
- ⑭ 아마 앞으로는 내가 전공을 택하던 때를 후회하고, 그때로 돌아가고 싶을 것이다. (K34)
- ⑮ 만약에 돌아간다면 어떻게 됐던 그녀에게 마음을 표현할 것이다. (K36)

위 ⑬~⑮와 같이 필자가 자신의 후회를 강조하기 위해 시간을 되돌려

과거로 다시 돌아가고 싶은 바람을 표현하는 구문을 실제 텍스트에서 확인할 수 있었다. 한국인 집단에서는 이처럼 상황을 가정하여 과거의 선택과는 다른 선택을 하고자 하는 바람으로 결말부를 서술한 경우가 많이 나타난 반면 외국인 집단의 경우 [그림 6]의 왼쪽 상단에서 하단으로 이어지는 ‘나, 후회하다, 가장, 소중하다, 우리’와 그 오른쪽 옆에 세로로 이어지는 ‘나중, 지금, 인생, 살다’의 묶음에서 볼 수 있다시피 후회를 극복하고 현재를 소중히 여기겠다는 내용과 ‘나중에 후회하지 않기 위해~’, ‘인생을 후회 없이 살고 싶으면~’과 같은 구문이 많이 사용된 것으로 보인다. 그리고 실제 학습자의 텍스트를 살펴본 결과 다음과 같은 결말부를 확인할 수 있었다.

- ⑩ 할아버지가 돌아가셔서 저를 세상에 후회하고 돌이킬 수 없는 일들이 많은 걸 알게 된 만큼 우리가 할 수 있는 일은 지금 가지고 있는 모든 것을 소중하게 생각하고 후회하지 않도록 하는 것이다. (F03)
- ⑪ 혼자 성공을 위해서 무정한 사람처럼 살면 나중에도 후회할 것이다. 그러므로 인생을 후회없이 살고 싶으면 소중한 것을 소중히 여기면 살아야한다. (F04)

외국인 집단에서는 ⑩, ⑪과 같이 앞으로의 다짐이나 당부를 메시지 전달 방식과 같이 서술된 경우가 자주 나타났으며 이는 앞서 언급한 과거의 상황을 가정하며 끝맺는 한국인 집단의 경우와 차이가 있다. 이와 관련하여 문법 사용면에서 역시 차이를 드러내는데 이에 대해서는 시제·상 부분에서 자세히 설명하도록 한다.

#### 4.2. 시제·상 분석

전술한 바와 같이 내러티브 텍스트는 경험을 기술하는 과정에서 시간에 관한 표현을 쓰기 때문에 필연적으로 시제·상과 같은 문법 요소가 자주 나타난다. 본 연구에서 분석한 텍스트에서도 과거시제의 사용이 많이 나타났음을 확인할 수 있었다. 다만 그 양상은 구조별로 두 집단 간의 차이가 있

였으며 과거시제 외 다른 문법 요소의 사용에서도 차이가 나타났다.

#### 4.2.1. 도입부

한국인 집단과 외국인 집단의 도입부에서 사용된 시제와 상의 빈도를 분석한 결과 다음과 같이 나타났다.

<표 7> 사건 도입부에 나타난 시제·상 빈도

순위	한국인			외국인			
	형태소		빈도	형태소		빈도	
1	ㄴ-/ETM	<과거시제전성어미>	54	ㄴ-/ETM	<과거시제전성어미>	40	
2	는/JX	<현재시제전성어미>	42	였/EP	<과거시제종결어미>	26	
3	였/EP	<과거시제종결어미>	30	는/JX	<현재시제전성어미>	26	
4	ㄹ-/ETM	<미래시제전성어미>	19	ㄹ-/ETM	<미래시제전성어미>	25	
5	던/ETM	<과거시제전성어미>	16	던/ETM	<과거시제전성어미>	10	
6	는다/EF	<현재시제종결어미>	3	고/EC	있.01/VX	<진행상>	6
7	을/ETM	것 <예정상>	3	겠/EP	<미래시제종결어미>	5	
8	고자/EC	<예정상>	3	을/ETM	것 <예정상>	5	
9	고/EC	있.01/VX	<진행상>	3	는다/EF	<현재시제종결어미>	3
10	어지/VV	<완료상>	2	다가/EC	<완료상>	2	

도입부에서는 서술되는 내용과 방법에 따라 현재시제 및 과거시제가 다르게 나타났는데, 한국인 집단의 경우 글을 쓰는 현재 시점에도 여전히 후회하거나 기억에 남는 사건을 기술하여 발화시 중심의 시제 사용이 많이 나타났으며 주로 주제를 소개하는 데 현재시제를 사용한 것으로 나타났다. 현재시제를 사용하여 도입 문장을 서술한 경우는 다음과 같다.

- ⑱ 나에게도 떠오르는 일이 몇 가지 있는데, 그중에서도 하나를 꼽자면 대학 입시 시절인 것 같다. (K10)
- ⑲ ‘내가 선택한 모든 것이 정답이다.’라는 좌우명답게 지금까지 내린 선택들 가운데 후회를 하는 일은 크게 없다. (K11)
- ⑳ 내가 지금까지 살아오면서 후회하는 일이 많지 않지만 가장 후회하는 일이 있다. (F06)

반면, 외국인 집단의 도입부에서 현재시제는 29회로 나타났는데 이는 한국인 집단의 45회에 비해 훨씬 낮은 빈도이다. 외국인 집단에서도 ㉔과 같이 현재시제를 사용하여 발화 당시까지도 사실인 내용을 기술하는 경우가 꽤 있었으나 아래 ㉕과 같이 후회하는 사건이 과거 어느 시점에 일어난 일이라는 것을 표현하기 위해 과거시제를 사용한 경우도 드물지 않게 발견할 수 있었다.

- ㉔ 나는 한국에 온 것은 후회한다. 왜냐하면 한국에 있는 동안 소중한 가족 두 분을 잃어버리게 되었기 때문이다. (F27)
- ㉕ 하지만 어떤 일들은 그 때 하지 않은 것을 후회한 적이 많다. (F31)
- ㉖ 어려서부터 지금까지 나는 가장 후회한 일이 바로 눈을 제대로 보호하지 못해 근시를 초래한 것이다. (F22)

㉔의 경우는 한국에 온 그 순간이 발화시 이전이기 때문에 과거시제가 사용되었으며 발화 시점까지 후회를 하고 있는 것으로 해석되어 종결부에서는 현재시제를 사용하는 것으로 볼 수 있다. 또한 ㉕와 같이 ‘-(으)ㄴ 적이 있다/많다’의 표현도 가끔 나타났으며 ㉖과 같이 현재시제 ‘-는’으로 실현되어야 자연스러움에도 과거시제를 사용하여 오류를 보여준 경우도 적지 않아 외국인 학습자가 시제 사용에 어려움을 겪는다는 것을 알 수 있었다.

#### 4.2.2. 사건 전개부

사건 전개부는 사건의 상황적 배경이나 사건 이해를 위해 필요한 정보를 제시하는 배경 소개 부분과 사건 서술 부분으로 이루어지는데 과거의 특정 사건을 기술해야 하는 주제의 특성상 과거 시제가 많이 나타날 것으로 예상 가능하다. 두 집단의 사건 전개부에 나타난 시제·상의 분석 결과는 다음과 같다.

<표 8> 사건 전개부에 나타난 시제-상 빈도

순위	한국인			외국인		
	형태소		빈도	형태소		빈도
1	있/EP	<과거시제종결어미>	276	있/EP	<과거시제종결어미>	242
2	는/JX	<현재시제전성어미>	209	ㄴ/ETM	<과거시제전성어미>	172
3	ㄴ/ETM	<과거시제전성어미>	183	는/JX	<현재시제전성어미>	172
4	던/ETM	<과거시제전성어미>	83	르/ETM	<미래시제전성어미>	103
5	르/ETM	<미래시제전성어미>	78	고/EC	있_01/VX <진행상>	27
6	고/EC	있_01/VX <진행상>	24	던/ETM	<과거시제전성어미>	20
7	어지/VV	<완료상>	16	겠/EP	<미래시제종결어미>	14
8	고/EC	말_03/VX <완료상>	9	어지/VV	<완료상>	11
9	을/ETM	것 <예정상>	9	고/EC	나_01/VX <완료상>	8
10	겠/EP	<미래시제종결어미>	8	자마자/EC	<완료상>	7

사건 전개부에서는 두 집단 모두 시제 중 과거시제 ‘있’이 가장 많이 사용되었으며 완료상을 제외한 예정상과 진행상은 낮은 빈도로 나타났다.

- ㉔ 나는 할머니 손에 **꿰**다. (K38)
- ㉕ 나는 대학교 1학년을 보내면서 동기들에 비해 많이 부족하다고 느꼈고, 한 계도 **분명했**다. (K01)
- ㉖ 나는 **어릴** 때부터 부모님과 함께 다른 도시에서 살기 때문에 할머니와 함께 있는 시간이 많이 **없**다. (F06)
- ㉗ 고등학교에 **다닐** 때 나는 공부 잘 안 하지만 성적이 나쁘지 않은 **편**이다. (F02)

㉔~㉗은 사건이 일어날 당시의 배경을 설명하고 있는 문장으로 한국인과 외국인 집단에서 대부분 과거시제를 사용하여 서술된 것을 볼 수 있었다. 분석 결과를 바탕으로 실제 텍스트를 살펴본 결과 외국인 집단의 경우 다른 구조부보다 사건전개부에서 시제 사용의 오류가 많이 나타난다는 것을 확인할 수 있었다. ㉖과 같은 경우 현재는 어렸을 때와 같은 삶을 살고 있지 않기 때문에 단절을 표현해야 하며 이를 위해서는 과거

시제를 사용하는 것이 자연스러운 것임에도 학습자는 현재시제를 사용하였다. 또한 ㉗과 같이 과거의 상황을 설명하면서도 과거시제를 사용하지 않고 절대시제를 사용하여 오류를 보인 경우도 흔하게 나타났다.

다른 특징으로는 과거시제 전성어미 ‘던’의 사용을 들 수 있는데 한국인 집단의 경우 83회 출현한 반면 외국인 집단에서는 20회로 나타나 그 차이가 매우 큰 것으로 나타났다.

- ㉘ 때는 중학생, 좋아하던 여학우가 있었다. (K15)
- ㉙ 익숙하고 계속해오던 일만 추구하느라 새롭게 도전하는 일을 망설였다. (K23)
- ㉚ 이후에 그를 볼 때마다 괜한 죄책감이 들었고, 그런 스스로가 한심해서 한 동안 우울감에 빠져 살았던 기억이 있다. (K18)
- ㉛ 두 살 차이인 동생도 대학에 들어간 후 엄마가 여유 좀 생겼다. 같이 보내던 3년 시간 내에 엄마가 직장 일 한 번도 얘기하지 않았다. (F28)

한국인 집단에서 사용된 ‘던’의 경우를 보면 ㉘~㉚과 같이 어떤 상황이 과거에 완료되지 않고 중단되었거나 현재와 단절된 사건을 회상하여 표현하기 위해 자주 쓰인 것으로 나타났는데 문장의 종결은 대부분 과거시제가 사용되었다. 반면 외국인 집단에서는 ‘던’의 사용이 매우 드물었으며 사용하더라도 ㉛과 같이 ‘있던’과 ‘던’을 정확하게 구별하여 사용하지 못하는 경우를 확인할 수 있었다. 외국인 학습자는 ‘던’의 사용 대신 단순 과거시제로 표현하는 경우가 많았는데 이는 한국어 중급 단계에서 해당 문법을 배웠을 당시 이미 초급에서 배운 과거시제의 사용이 익숙해져 새로운 문법을 잘 활용하지 못한 채 고급 단계까지 이르렀거나 ‘던’이 지닌 회상의 기능 등을 잘 인식하지 못하고 있기 때문인 것으로 보인다.

또한 두 집단은 완료상의 사용에서도 차이가 나타났다.

- ㉜ 나는 사람들 앞에서 이야기하는 것이 두렵다는 이유로 면접을 이틀 정도 앞둔 시점에서 결국 그 시험을 포기해 버리고 말았다. (K10)

- ㉓ 그렇게 우리 관계는 끝이 나고 말았다. (K35)
- ㉔ 친한 친구가 알고 나서 나에게 화가 많이 났다. (F13)
- ㉕ 집안에 가장 중시하는 오빠의 결혼식을 끝나고 나서 서울에 다시 가는 날도 되었다. (F24)

한국인 집단은 ㉒, ㉓과 같이 완료상 ‘-고 말다’를 과거시제 형태소를 첨가하여 ‘-고 말았다’와 같이 표현한 경우가 9회 나타난 반면 외국인 집단에서는 해당 표현이 단 1회 나타났다. 한국인 집단에 비해 외국인 집단에서 많이 등장한 완료상 ‘-고 나다’가 있었는데 주로 ㉔, ㉕와 같이 일의 순서를 나타내는 기능으로 사용하였다. 외국인 학습자들은 다양한 상과 시제를 결합하고 사용하는 것에 미숙해 보이며 이에 대해서는 좀 더 면밀한 연구가 필요한 것으로 보인다.

### 4.2.3. 결말부

결말부는 사건이 현재 자신의 상황에 영향을 주거나 그로 인해 결심한 내용이 주로 기술되며 내러티브를 끝맺는 기능을 한다. 결말부에 나타난 시제·상 분석 결과는 다음과 같다.

<표 9> 결말부에 나타난 시제·상 빈도

순위	한국인			외국인		
	형태소	빈도		형태소	빈도	
1	는/JX	<현재시제전성어미>	63	ㄴ/ETM	<과거시제전성어미>	45
2	ㄴ/ETM	<과거시제전성어미>	56	ㄹ/ETM	<미래시제전성어미>	44
3	ㄹ/ETM	<미래시제전성어미>	49	는/JX	<현재시제전성어미>	37
4	었/EP	<과거시제종결어미>	41	었/EP	<과거시제종결어미>	29
5	던/ETM	<과거시제전성어미>	29	고/EC	있_01/VX <진행상>	14
6	고/EC	있_01/VX <진행상>	8	도록/EC	<예정상>	8
7	을/ETM	것 <예정상>	8	겠/EP	<미래시제종결어미>	7
8	겠/EP	<미래시제종결어미>	5	던/ETM	<과거시제전성어미>	5
9	고자/EC	<예정상>	4	는다/EF	<현재시제종결어미>	3
10	도록/EC	<예정상>	3	어/EC	있_01/VX <완료상>	2

결말부에서 사용된 시제의 경우 현재시제와 과거시제 ‘던’의 차이가 확연히 나타났다. 먼저 현재시제 ‘는’의 경우 한국인 집단에서는 63회, 외국인 집단에서는 37회 나타나 큰 차이를 보였으며 과거시제 ‘던’의 경우 역시 한국인 집단에서 29회, 외국인 집단에서 5회로 나타나 대조적인 결과를 보였다<sup>10)</sup>. 이는 한국인의 경우 과거의 후회했던 순간이나 사건을 ‘엿던’을 사용하여 정리하며 현재에도 느끼는 감정을 진술하면서 끝맺는 경우가 많은 반면, 외국인의 경우 후회는 이미 지나간 일이므로 현재 또는 미래에 해야 할 일을 강조하거나 결심을 서술하면서 글을 마치는 경우가 많았기 때문이다.

- ㉞ 2년이라는 긴 시간동안 벽 안에 평범한 사람을 가두고 외면했던 그 순간들이 내가 인생에서 가장 후회하는 순간들이다. (K31)
- ㉟ 결과가 두려워서 하고 싶은 마음을 숨기고 현실에 안주하려 했던 그 시절의 모습이 내가 가장 후회하는 일이다. (K36)
- ㊱ 그러나 알바를 못 해 본 것은 이미 엿질러진 물이니 이제 와서 후회해도 소용 없다. 앞으로 후회할 일을 저지르지 않도록 무슨 일을 하기 전에 잘 생각해둬야 한다. (F01)
- ㊲ 과거를 되돌리는 방법은 없다. 아무리 후회해도 슬퍼도 앞길을 쪽 걸어야 하고 후회 없이 살고 싶다면 가지고 있는 것이든 곁에 있는 사람이든 소중히 여기는 것이 중요하다. (F27)

한국인 집단은 ㉞, ㉟과 같이 사건에 대한 회상과 함께 다시 주제를 언급하는 방식으로 서술하는 경우가 많이 나타났다. 이는 사건전개부에서 과거 사건과 그에 대한 평가, 감상 등을 기술한 후 결말부에서 현재 시점

10) 이에 대한 원인은 허명나·박소영(2021)을 참고할 수 있다. 중국어권 중·고급 학습자 140명을 대상으로 ‘-은/던/엿던’의 사용 양상에 대한 설문 조사 결과 응답자들은 전체적으로 상대시제 이해 부족 및 ‘-던/엿던’ 사용을 회피하는 것으로 나타났으며 ‘-은/던/엿던’의 정확한 의미 및 그 차이를 충분히 인지하지 못하고 있음을 드러냈다.



으로 돌아와 그때의 일을 다시 언급하며 정리하는 방식으로 서술함으로써 주제를 명료하게 하고 현재 글을 읽는 독자에게 필자의 과거 경험에 대해 교류의 감정을 느끼게 하는 효과를 가질 수 있다.

- ㉔ 나도 가족들이 내가 시험 준비를 열심히 했으면 좋겠다는 것을 알았다. 다시는 이런 후회하는 일이 없기를 바란다.
- ㉕ 이번에 겪은 일을 통해 자신이 잘못하고 있다는 점을 깨닫고 앞으로도 다른 사람의 입장에서 더 잘 생각하게 되기를 바란다. 마지막으로는 부모님에게 나의 변화를 실감하고 성숙한 사람이 되었다는 것을 실감했으면 좋겠다. (F36)

한국인 집단에서는 나타나지 않은 구문이 외국인 집단에서만 나타난 경우가 있었는데 ‘-기를 바란다’와 ‘-(으)면 좋겠다’와 같은 표현으로 ㉔, ㉕과 같이 학습자들은 앞으로의 다짐이나 소망을 기술하기 위해 해당 표현을 사용하였다. 비슷한 맥락의 문장을 서술할 경우 한국인 집단에서는 ‘-(으)르 것이다’, ‘-겠-’의 양태 표현이 많이 쓰인 반면 외국인은 주관성이 더 강하게 느껴지는 ‘-(으)면 좋겠다’가 자주 쓰였다는 점에서 볼 때, 또한 ‘-기를 바라다’의 경우 대상이 자신 스스로에게 잘 쓰지 않음에도 외국인들의 텍스트에서 나타났다는 것은 외국인 학습자가 배운 표현의 제약 사항에 대해 정확히 알지 못하며 양태성에 대해서도 잘 인식하지 못하고 있다는 것을 짐작할 수 있다. 아울러 한국어 초·중급 단계에서 배운 표현을 대체할 다른 표현도 능숙하게 사용하지 못한다는 것을 알 수 있었다.

## 5. 결론 및 제언

본 연구는 한국어 학습자들이 생산한 경험적 내러티브 텍스트의 구조를 분석하여 그 안에서 나타나는 어휘와 시제·상의 사용 양상을 한국인

모어 화자 집단과 비교하여 살피고자 하였다. 텍스트 구조에 따라 나타난 언어적 표현의 특징을 살펴보는 것은 필자의 어휘 및 문법 습득의 정도와 양상을 파악할 수 있을 뿐만 아니라 한국인 모어 화자와의 차이를 발견하여 오류의 원인을 탐색하는 데에 도움이 된다. 또한 본 연구에서 핵심 어휘와 문법 요소를 분석하는 도구로 활용된 텍스트 마이닝과 언어 네트워크 기법은 정성적인 연구 방법을 통해서서는 확인하기 어려웠던 언어적 특징을 전산화하여 객관적으로 보여줄 수 있다는 장점이 있다.

본 연구에서 ‘인생에서 가장 후회하는 일’이라는 주제로 한국인 모어 화자와 외국인 학습자의 내러티브 텍스트를 수집하여 분석한 결과, 구조별 출현 양상은 비슷하게 나타났다. 두 집단의 필자들은 대부분 도입부-사건전개부-결말부 구조에 따라 글을 완성하였으며 각 텍스트의 문장 및 문단 수, 총 어절 수 및 각 구조부의 어절 수에서도 큰 차이를 보이지 않았다.

각 구조에 나타난 핵심 어휘와 시제·상의 사용 양상 분석 결과에서는 부분적으로 크고 작은 차이가 나타났다. 먼저 핵심 어휘와 네트워크에 나타난 어휘 사용 양상을 분석한 결과, 도입부에서 주제를 소개하기 위해 한국인 모어 화자는 누구나 느낄 만한 보편적인 내용으로 비교적 객관성을 지닌 문장으로 도입하는 것을 선호하는 반면 외국인 학습자는 주로 자신의 감정이나 주관적인 견해를 직접적으로 표현하며 글을 시작한다는 점을 알 수 있었다. 외국인 학습자의 도입부에서 특히 자주 출현한 어휘는 1인칭을 나타내는 명사였는데 이는 실제 텍스트를 살펴본 결과 학습자가 1인칭 대명사를 과잉 사용하고 있는 것으로 확인되었다. 사건전개부에서는 두 집단 모두 학업 관련 사건을 다뤘다는 공통점을 발견할 수 있었는데 외국인 학습자의 경우 가족과의 일화나 가족과 함께 시간을 보내지 못하는 상황을 소재로 서술한 경우 역시 많았다는 점이 특징이었다. 결말부에서는 한국인 집단의 경우 상황을 가정하여 과거의 선택과는 다른 선택을 하고자 하는 바람으로 서술한 경우가 많이 나타난 반면 외국인 한국어 학습자는 후회를 극복하고 현재를 소중히 여기겠다는 내용과

미래 전망과 같은 내용으로 마무리되는 경우가 많았다.

다음으로 시제·상의 사용 양상 분석을 통해 두 집단 모두 도입부와 사전전개부에서 과거 시제의 사용 빈도가 가장 높다는 것을 알 수 있었다. 다만, 도입부에서 한국인 집단은 외국인 집단에 비해 현재시제 사용 빈도가 높게 나타났으며 전개부에서는 과거시제 ‘던’과 완료상 ‘고 말’이 더 자주 출현한 것으로 나타났다. 또한 결말부에서는 과거시제 ‘던’과 현재시제 사용 빈도가 한국인 집단에서 더 높게 나타났다. 외국인 학습자의 경우 ‘던’과 ‘있던’의 사용이 한국인 집단에 비해 훨씬 낮은 빈도로 나타났는데 이를 통해 학습자가 다양한 시제를 사용하지 못하고 단순과거시제로만 사건의 내용을 서술한다는 점을 알 수 있었다. 또한 완료상을 비롯한 다양한 상과 양태성을 지닌 문법 요소에 대해서도 능숙하게 활용하지 못한다는 점이 확인되어 이와 관련된 학습이 요구된다는 것을 알 수 있었다.

본 연구는 텍스트 마이닝을 활용하여 외국인이 실제 작성한 텍스트의 언어적 특징을 한국인 모어 화자 비교 분석했다는 점에서 의의가 있다. 다만 여전히 정교화되지 못한 기술력은 본 연구가 앞으로 해결해 나갈 점이라고 생각한다. 언어 네트워크에 대한 분석 방법이 좀 더 정밀해진다면 시제·상 뿐만 아니라 다양한 문법 요소에 관한 분석도 가능해질 것이라고 본다. 이러한 연구는 후속 연구를 통해 진행하고자 한다.

---

\* 이 논문은 2022년 10월 10일에 접수하여 2022년 11월 1일에 심사를 완료하고 2022년 11월 7일 게재를 확정함.

<참고 문헌>

- 강은실(2007), 「서사 텍스트의 거시구조에 대한 이론적 검토와 교육적 활용 방안 연구」, 『국어교육연구』 40, 국어교육학회, 1-34.
- 남기심·고영근·유현경·최형용(2019), 『새로 쓴 표준 국어문법론』, 한국문화사.
- 심은지(2011), 「한국어 학습자 작문에 나타난 시제상 습득 연구」, 이화여자대학교 석사학위논문.
- 오지혜(2008), 「이야기 문법을 활용한 한국어 고급 학습자의 이야기 쓰기 교육 연구」, 『작문연구』 7, 한국작문학회, 275-305.
- 오지혜(2016), 「한국어교육에서 내러티브 텍스트 연구」, 『한국언어문화교육학회 학술발표논문집』, 한국언어문화교육학회, 99-111.
- 유민애(2017), 「맥락 중심의 한국어 담화문법 교육 연구」, 서울대학교 박사학위논문.
- 조국현(2013), 「인터넷 일상 서사 텍스트의 구조와 기능」, 『독어교육』 56, 한국독어독문학교육학회, 259-287.
- 조은(2015), 「한국어 학습자의 내러티브 텍스트 분석」, 이화여자대학교 석사학위논문.
- 채윤마·길호현(2020), 「언어 네트워크 분석을 활용한 한국어 쓰기 교육 연구」, 『이중언어학』 31, 이중언어학회, 173-200.
- 한명나·박소영(2021), 「중국어권 한국어 학습자를 위한 과거시제 관형사형 어미 ‘-은/던/있던’ 교육 연구」, 『이중언어학』 84, 이중언어학회, 393-416.
- 한송화(2015), 「한국어 학습자의 경험적 내러티브에 나타난 구조적 특성과 언어 사용 양상」, 『국어교육』 148, 한국어교육학회, 511-546.
- Beaugrande, R. A., & Dressler, W. U.(1981), *Introduction to text linguistics*, London: Longman.
- Labov, W.(1972), *Language in the inner city: Studies in the Black English vernacular*, Philadelphia: University of Pennsylvania Press.

## <국문 초록>

본 연구는 외국인 한국어 학습자가 생산한 내러티브 텍스트의 언어적 표현에 대한 사용 양상을 한국어 모어 화자의 텍스트와 비교하여 살펴보고, 객관적인 방법으로 그 특징을 밝히고자 하였다. 이를 위해 기존에 이루어졌던 분석 방법과 차별화를 두어 텍스트 마이닝의 언어 네트워크 기법을 활용하였다. 통계학적으로 유의미하다고 판단되는 어휘와 문법적 요소를 추출한 후, 텍스트 구조에 따라 어휘와 시제·상의 사용이 어떻게 다르게 나타나는지 살펴보았다. 두 집단은 대부분 도입부, 사건 전개부, 결말부 구조에 따라 글을 전개하였으며 큰 차이를 보이지 않았다. 구조별 핵심 어휘와 어휘 간 관계성을 살펴본 결과 도입부, 사건전개부, 결말부에서 두 집단은 주제에 따른 내용의 차이를 보였으며 외국인 학습자의 경우 도입부에서 1인칭 대명사를 과도하게 사용하는 특징이 나타났다. 또한 각 구조에 나타난 시제·상의 사용 양상을 살펴본 결과 외국인 학습자는 한국인 모어 화자에 비해 시제와 상, 그리고 양태성을 지닌 문법 요소에 대해서도 다양하게 활용하지 못한다는 점이 확인되었다. 이는 학습자가 초급에서 배운 문법 사용에 익숙해져 있으며 시제와 상과 관련한 습득은 아직 불완전한 것으로 보여 이에 대한 학습이 요구된다는 것을 알 수 있었다. 본 연구는 정성적인 연구 방법을 통해서는 확인하기 어려웠던 언어적 특징을 전산화하여 명시적으로 파악하고 결과를 도출하려고 했다는 점에서 의의가 있다.

### □ 주제어

내러티브 텍스트, 텍스트 마이닝, 언어 네트워크 분석,  
외국인 한국어 학습자, 한국어 쓰기 교육

<Abstract>

## Keywords and Grammar Usage Patterns of Korean Learners Using Text Mining

- Focused on narrative text -

Chae Yun-mi · Kil Ho-hyun · Wang Yan

This study aims to examine the usage patterns of linguistic expressions in narrative texts produced by foreign Korean learners compared with the text of Korean native speakers, and to reveal their characteristics objectively. To this end, the semantic network technique of text mining was used to differentiate it from the existing analysis methods. After extracting the words and grammatical elements considered statistically significant, we examined how the use of word, tense, and aspect differed according to the text structure. Korean native speakers and foreign Korean learners mostly developed their writings according to the structure of the introduction, events, and conclusion, and there was no significant difference. As a result of examining the relationship between key words and words for each structure, the two groups showed differences in content according to the topic in the introduction, events, and conclusion. Foreign learners displayed a characteristic of excessive use of first-person pronouns. In addition, as a result of examining the usage patterns of tenses and expressions in each structure, it was confirmed that foreign learners were unable to utilize the grammatical elements with tenses, aspect,

and modalities as variedly as native Korean speakers did. This shows that the learner is accustomed to using grammar learned at the beginner level, and the acquisition related to tense and phase is still incomplete; hence, learning is required. This study is meaningful in that it attempted to computerize and explicitly identify linguistic features that were difficult to confirm through qualitative research methods to derive at results.

□ **Keywords**

narrative text, text-mining, semantic network analysis,  
foreign Korean learner, Korean writing education