

http://dx.doi.org/10.17703/JCCT.2021.7.4.313

JCCT 2021-11-38

문학 텍스트의 머신러닝 활용방안 연구  
- 화자 지시어 분석을 위한 규칙 선별을 중심으로 -

**A Study on the Application of Machine Learning in Literary Texts  
- Focusing on Rule Selection for Speaker Directive Analysis -**

권경아\*, 고일주\*\*, 이인성\*\*\*

**Kyoungah Kwon\*, Ilju Ko\*\*, Insung Lee\*\*\***

**요약** 본 연구는 문학 텍스트를 활용한 머신러닝 기반 가상 캐릭터(virtual character) 구현을 위해 텍스트 내의 화자 지시어가 지시하는 화자를 판별할 수 있는 규칙을 제안하는 것을 목적으로 한다. 선행 연구에서, 본 연구자는 문학 텍스트를 기계 학습에 적용할 때, 별칭, 별명, 대명사와 같은 화자 지시어들이 특정한 분석 규칙 없이 기계가 화자를 제대로 파악하지 못하여 학습을 제대로 수행할 수 없다는 점을 발견하였다. 본 연구는 이를 해결하는 방법으로 ‘화자 지시어(대명사 포함)가 지시하는 화자를 찾는 9가지 규칙’을 소개한다: 위치, 거리, 대명사, 가주어/진주어, 인용문, 화자 수, 등장인물 외 지시, 복합 단어 지시, 화자명 분산이 그것이다. 문학 텍스트 내의 등장인물을 가상 캐릭터로 활용하기 위해서는 기계가 이해할 수 있는 방식으로 학습 텍스트를 제공해야 한다. 본 연구자는 본 논문을 통해 제안한 화자 찾기 규칙이 문학 텍스트를 머신러닝에 활용할 때 발생할 수 있는 시행착오를 줄이고, 원활한 학습을 수행하게 하여 질적으로 우수한 학습 결과를 산출할 수 있게 해 줄 것으로 기대한다.

**주요어** : 문학 텍스트, 화자 지시어, 영어 대명사, 대응화, 머신 러닝

**Abstract** The purpose of this study is to propose rules that can identify the speaker referred by the speaker directive in the text for the realization of a machine learning-based virtual character using a literary text. Through previous studies, we found that when applying literary texts to machine learning, the machine did not properly discriminate the speaker without any specific rules for the analysis of speaker directives such as other names, nicknames, pronouns, and so on. As a way to solve this problem, this study proposes ‘nine rules for finding a speaker indicated by speaker directives (including pronouns)’: location, distance, pronouns, preparatory subject/preparatory object, quotations, number of speakers, non-characters directives, word compound form, dispersion of speaker names. In order to utilize characters within a literary text as virtual ones, the learning text must be presented in a machine-comprehensible way. We expect that the rules suggested in this study will reduce trial and error that may occur when using literary texts for machine learning, and enable smooth learning to produce qualitatively excellent learning results.

**Key words** : Literature, Speaker-directed Terminology, English Pronoun, Pronominalization, Machine Learning

\*정회원, 숭실대학교 글로벌미디어학부 강사 (제1저자)  
\*\*정회원, 숭실대학교 글로벌미디어학부 교수 (제2저자)  
\*\*\*정회원, 숭실대학교 영어영문학과 교수 (교신저자)  
접수일: 2021년 9월 30일, 수정완료일: 2021년 10월 10일  
게재확정일: 2021년 10월 18일

Received: September 30, 2021 / Revised: October 10, 2021  
Accepted: October 18, 2021  
\*Corresponding Author: peace@ssu.ac.kr  
Dept. of English Language and Literature, Soongsil Univ.

## 1. 서론

메타버스(Metaverse) 환경의 대두는 메타버스 기술 뿐만 아니라 다양한 개념의 공간 및 캐릭터에 관한 연구의 필요성을 함께 제기한다. 게임과 커뮤니티 서비스를 결합한 로블록스(roblox)와 포트나이트(Fortnite)를 비롯해서, 페이스북의 소셜 VR 플랫폼, 증강현실 아바타 서비스 제페토(Zepeto) 등 현재 Pre-Metaverse로 불리는 서비스 대부분은 실시간 소셜 커뮤니티를 활성화 하려는 목적으로 현실 공간에 가깝게 구축하거나 게임 환경을 차용하는 방식으로, 기존 VR 환경의 연장선상에서 메타버스 환경의 구축 방향을 모색하고 있다.

문학 작품은 작가가 창조한 시공간과 그곳에서 행하는 다양한 캐릭터들을 텍스트의 형식으로 담고 있다. 오랜 역사 속에서 방대하게 생산된 문학 텍스트 내의 시공간을 가상 환경을 비롯한 메타버스 환경 구축에 적극적으로 활용한다면, 현재의 게임 및 비즈니스 중심의 가상 환경을 문화 원형이 재현된 공간으로 그 영역을 넓힐 수 있다. 더해서, AI 기술을 기반으로 문학 텍스트 내의 인공지능을 가상 현실 혹은 메타버스 환경의 캐릭터로 재구축하는 연구는 텍스트적 캐릭터의 매체적 전이를 통해 캐릭터성의 확장적 재현도 가능하게 한다.

문학 텍스트 내의 등장인물을 활용해서 가상 현실 혹은 메타버스 환경을 구축할 때 머신러닝을 도입해야 하는 또 다른 이유는 분석해야 할 텍스트의 분량이 적지 않고 동일한 작업을 반복해야 하는 특성 때문이다. 문학 텍스트에는 한 작가가 기고한 여러 편의 소설이나 시리즈물도 존재한다. 각 작가마다 고유의 표현 형식과 이야기 전개 방식을 가진다는 점에서, 일반화된 텍스트 분석 방법 없이 개별 작가들의 텍스트를 분석하는 것은 불필요한 반복작업을 초래할 가능성을 키운다. 한 작가의 대표 작품을 기준으로 문학 텍스트에 특화된 공통 규칙을 사전에 정의하고, 이것을 기계 학습에 대입하여 결과를 확인한 후 동일 작가의 다른 작품이나 다른 작가들로 분석을 확대하면 중복 작업을 피할 수 있다. 문학 작가들이 문학이라는 테두리 안에서 ‘유사하면서도 다른’ 화법을 사용한다는 점을 참작하면, 본 연구는 문학 작품 전반에 적용할 수 있는 ‘공통 규칙’뿐만 아니라 개별 작가들의 스타일이 반영된 ‘특수 규칙’도 함께 확인할 기회가 된다. 지금까지 작가별로 화자를 지시하는 용어들의 사용 용법이나 스타일의 차이를 연구한 사례

를 찾을 수 없는 것은 연구의 목적이 작가들을 구별하기 위해서가 아니라, 단순히 해당 작가의 특징을 찾아서 글을 쓰기 위한 참조 용도로 사용했을 가능성이 크기 때문이다. 따라서, 머신러닝에 문학 텍스트를 활용하면 분석 작업의 효율성을 높일 뿐만 아니라, 다른 관점에서 텍스트 내의 화자 지시어 용법과 작가 스타일 연구를 바라볼 수 있게 하여 문학 텍스트 자체의 연구 영역의 확장도 함께 가져올 수 있다.

본 연구자는 이전 연구를 통해 머신러닝을 활용하여 문학 텍스트 내의 캐릭터 추출 작업을 시도한 바 있다. “문학 텍스트를 활용한 머신 러닝 언어모델 구현” [1]은 Sequence to Sequence, Transformer, GPT-2 등 대표적인 대화형 기계 학습 모델에 문학 텍스트를 적용한 결과를 제시한, 본 논문의 선행 연구다. 이 실험에서, 본 연구자는 문학 텍스트를 기계 학습에 도입하기 위해서는 사전에 세 가지 텍스트 보정 작업—대명사 처리, 대화쌍 늘리기, 데이터 증폭—을 필수적으로 수행해야 한다는 것을 제안하였다. 보정 작업을 거치지 않은 순수 문학 텍스트로는 기계가 학습을 제대로 수행할 수 없었기 때문이다. 문학 텍스트는 일상 대화문이나 시사 상식처럼 불특정 다수를 대상으로 하는 대화문으로 구성되지 않거나, 질문에 대한 답변이 분명하게 제시되지 않는 경우가 많다. [1] 특히, 가장 큰 문제로 대두된 것은 화자를 지시할 때 화자명을 사용하지 않고 ‘대명사’를 이용한다는 점이었다. 일반적으로 영어의 문어체적 표현은 화자의 이름을 반복사용하지 않고 처음 그 이름을 표기한 이후에는 해당 화자의 성별을 반영한 대명사로 대체하여 표현한다. 이 때문에, 기계가 대명사가 지시하는 화자를 분별하지 못하여 학습을 제대로 수행하지 못했고, 이를 해결하기 위해 본 연구자는 대명사뿐만 아니라, 화자의 외모나 성격을 반영한 ‘별명’과 화자를 달리 부르는 이름인 ‘별칭,’ 그리고 Mr.와 같이 화자의 성별이나 직위 등을 표현하는 ‘의존명사’ 등을 수작업으로 진행해야만 했다.

본 연구는 기계 학습을 목적으로 문학 텍스트 내의 화자 지시어가 지시하는 화자를 판별할 수 있는 규칙을 제안하는 것을 목적한다. 이전 연구에서 특히 수작업으로 진행해야만 했던, ‘대명사가 지시하는 화자명을 기계가 직접 판별할 수 있는’ 지시 패턴을 찾는 것이 그것이다. 이를 위해, 본 연구는 본 논문을 통해 ‘화자 지시어(대명사 포함)가 지시하는 화자를 찾는 9가지 규칙’을 제안한다.

연구 방법은 다음과 같다. 먼저, 분석 텍스트의 제1장을 샘플링 범위로 정하고, 텍스트에 포함된 화자 지시어 중에서 별명, 별칭에 해당하는 용어 목록을 만들어 분리하고, 남은 화자 지시어 중에서 사람을 지시하는 용어와 그렇지 않은 용어를 다시 분리한 후, 사람을 지시하는 화자 지시어를 다시 화자명과 화자 지시어 사이의 '거리,' 문장 내에서 화자 지시어의 '위치,' 해당 화자 지시어가 '인용문' 안에 포함되어 있는지 등, 화자 지시어의 사용 용법에 따라 다시 재분류한다. 이렇게 찾은 규칙들을 텍스트 전체로 확대 분석해서 오차 범위와 출현 빈도수를 확인하고, 마지막으로 추려진 용법을 토대로 문학 텍스트에서 화자 지시어를 고유명사로 바꿀 수 있는 '기본 규칙(Basic Rule)'을 정리한다.

본 연구는 분석 텍스트로 이전 연구에서 학습 데이터로 활용한 C. S. 루이스의 『나니아 연대기』(*The Chronicles of Narnia*) (1950~1956) 시리즈 중 제1권 『사자와 마녀와 옷장』을 선택하였다. [2] 본 연구가 문학 텍스트 내의 캐릭터를 가상 캐릭터로 구현하는 것에 목적을 두기 때문에, 주요 분석 대상은 텍스트 안의 화자 지시어 중에서도 '등장인물을 지시하는 용어,' 즉 '누구'를 지칭하는 용어로 제한하고, This, That과 같이 화자와 관련이 없는 지시어들은 제외하였다.

화자명, 즉 고유명사(names)에 대응하는 용어로는 별칭(another name), 별명(nickname), 동격명사(appositives), 특정 명사를 가리키는 명사구(definite noun phrases), 대명사(pronouns), 서술 명사(predicate nominals) 등이 있다. 별명, 별칭의 경우 화자를 직접 지시하기 때문에 학습을 필요로 하지 않는다. 쉽표 혹은 that 절로 표현되는 동격명사나 명사구는 그 안에 포함된 대명사가 지시하는 화자를 찾는 것으로 포괄하고, 기능 동사와 결합해서 주로 사용되어 명사적 특성과 술어적 특성을 동시에 가지는 서술 명사는 화자의 행위를 지시하는 동사와 함께 연구되어야 할 대상으로 보고 이번 화자 지시어 연구에서는 제외하였다.

화자 지시어 중에는 문맥에 대한 이해 없이는 지시 대상이 누구인지를 분명히 파악할 수 없는 단어들이 있다. 이런 지시어들은 학습자가 텍스트의 내용을 충분히 파악한 상태에서 문장에 의미론적(semantic)으로 접근해야 화자 지시어의 화자를 찾을 수 있다. 이런 유형의 지시어 분석은 분석의 완결성이라는 측면에서는 필요한 작업이지만, 문장 자체가 전문가도 해석하기 쉽지

않은 경우가 많아서 이번과 같은 초기 단계의 작업에는 적절하지 않다 판단하고 분석에서 제외하였다.

## II. 관련 연구

텍스트 내에서 발화자를 지시하는 '화자 지시어'에는 대명사, 별명, 의존명사, 대명사와 명사가 결합된 경우, 가주어 it에 대응하는 for, that 절 등의 진주어가 포함된다. 이 중 본 연구의 분석 텍스트에서 가장 많이 출현한 품사는 대명사다. ([표-4] 참조) 대명사는 본래도 사람이나 사물의 이름을 대신 나타내기 위해 만들어진 용법이다. 대명사에는 인칭 대명사와 지시 대명사, 관계 대명사 등이 있고, 인칭 대명사는 또 주격, 목적격, 소유격, 재귀 대명사 등으로 나뉜다. 실제로 언어학 및 정보처리, 딥 러닝 등 담화 분석 혹은 자연어 처리 연구에서도 대명사를 분석하는 작업이 가장 많이 수행되고 있다. 이에 따라, 본 연구는 대명사 연구를 본 연구의 주요 관련 연구로 보고 언어학 및 정보처리/딥 러닝에서의 연구를 조사하였다.

### 1. 언어학에서 대명사 연구 관련

대명사에 매칭되는 화자의 명칭, 즉 고유명사를 찾기 위한 연구는 다음과 같이 3가지로 요약된다. 첫 번째 방법은 전통적인 언어 지식(linguistic knowledge)과 도메인 지식(domain knowledge)을 이용하는 것이다. 이것은 형태소와 구문, 의미 및 담화 분석을 통해 언어가 가진 품사와 구문, 의미정보 등을 파악해서 이를 기반으로 대명사와 고유명사 사이의 상호 조응 관계를 파악하여 결정하는 방식이다. 두 번째 방식은 첫 번째 방식을 보완하기 위한 knowledge-poor approach 방법으로 형태소 분석이나 품사 태깅 결과를 이용하거나, 부분 구문 분석(partial parsing)을 기반으로 한다. 마지막 방식은 말뭉치에 기반한 통계적 기법으로 대용어 관계가 미리 설정된 말뭉치에서 관계 파악을 위한 패턴 정보를 학습하는 방식이다. [3][4][5]

언어학에서 대명사 관련 연구는 전통적인 언어 지식과 도메인 지식을 이용하는 방식으로, 품사 정보, 구문 정보, 의미정보 등을 이용해서 상호 조응 관계를 조사하는 방식이다. Hobbs(1976)는 대명사가 발견된 위치에서 앞부분의 명사구들에 대해 순서대로 성(gender), 수(number)와 같은 제약적 선택 조건을 적용하여 처음으로

로 이 조건을 만족하는 명사구를 선행사로 결정하는 알고리즘을 제안한다. [6] CogNIAC에서는 영어 대명사의 선행사를 결정하는 6가지 경험 규칙을 제시한다. 1. 현재 분석 중인 단어의 앞부분에서 선행사라고 판단되는 단어가 유일하면 이를 선택한다. 2. 현재 분석 중인 단어가 재귀 대명사이면 현재 문장에서 가장 가까운 선행사 후보를 선택한다. 3. 현재의 문장과 바로 이전문장에서 선행사라고 판단된 단어가 유일할 경우 이를 선택한다. 4. 현재 분석 중인 단어가 소유대명사이면 exact string match인 단어를 선택한다. 5. 현재 문장에서 선행사라고 판단된 단어가 유일하면 이를 선택한다. 6. 이전문장에서 선행사로 찾은 단어와 현재 분석 중인 단어가 격이 같은 경우에는 선택한다. [7]

Hobbs와 CogNIAC의 연구는 본 연구에서 추출한 규칙이 언어학에서 연구되는 방향과 일맥 하다는 것을 보여준다. 특히, CogNIAC의 6가지 경험 규칙은 본 연구에서 선별한 규칙과 함께 머신 러닝의 화자 지시어 찾기 학습에 적용할 경우, 화자가 모호한 화자지시어를 기계가 분별하는 데 유의미한 방향성을 제시해 줄 수 있을 것으로 보인다.

## 2. 정보처리/딥 러닝 관련

정보처리 및 딥 러닝 관련 연구에서도 기계 학습을 위한 대명사 처리 연구가 시도된다. 대명사 상호참조 해결을 위한 연구에서는 RNN(Recurrent Neural Network)을 기반으로 어텐션 메커니즘(Attention mechanism)을 이용하여 "입력 시퀀스에 대응되는 위치들의 리스트를 출력하는 딥 러닝 모델"을 제안한다. [8] 문서 내에 등장하는 대명사가 어떤 선행사를 가리키고 있는지를 밝히기 위해 RNN의 인코더-디코더 네트워크에 각각 형태소 정보(위치를 포함하여)와 대명사를 입력하여 테스트한 점이 본 연구에서 향후 시도할 기계 학습 모델을 구성하는 데 참조가 되었다. 특히, 멘션 페어 모델에서 선행사(본 연구에서는 화자에 해당하는)와 대명사가 멀리 떨어져 있는 경우에 알고리즘이 주변 문맥만으로 판단하게끔 모델이 설계되어 오류가 자주 발생하는 점에 대해서, [9] 위 연구에서 이를 보완하는 방식으로 RNN의 확장형 모델인 포인터 네트워크를 이용하면 문맥 정보를 활용해서 추적할 수 있다고 제안하여, 본 연구에서 예외 규칙으로 선별된 문체에 접근하는 데에 많은 참조가 되었다.

## III. 텍스트 분석

문학 텍스트에서는 페르소나를 추출할 때 일반적으로 형용사와 부사, 동사를 가장 많이 참조한다. 형용사는 등장인물의 크기(long, big, tall, etc.), 모양(round, square, etc.), 칼라(yellow, blue, etc) 혹은 감정 상태(sad, glad, etc.)등을 주로 표현한다. 부사는 동사의 뒤에 위치해서 수식하는 동사의 반복 빈도나 시간적 간격(always, sometimes, often, usually, etc.), 행위 혹은 상태의 정도(completely, slowly, etc.)를 나타내며, 형용사 혹은 다른 부사를 강조해서 정도의 차이를 드러내기도 한다(very, enough, early, pretty, etc.). 동사는 서술 명사와 함께 화자의 행동적 특징을 주로 표현한다. 저자는 부사와 동사를 결합해서 화자의 행위성의 정도, 가령 화자가 능동적인 타입인지 수동적인 타입인지를 전달할 수 있다. 하지만, 형용사, 부사 및 동사 같은 품사들은 등장인물의 특징을 '지시(indicate)'하는 것이 그 인물의 페르소나 자체를 정의하는 용어들이 아니다. 등장인물의 성격은 그것이 작가의 머릿속에서든 특정 문서를 통해서든 품사를 이용해서 글로 서술되기 전에 '미리-결정(pre-decision)'되기 때문이다. 형용사, 부사, 동사가 이러한 '성격-원형(personality antetype)'으로부터 유추된다는 점을 고려하면, 이러한 품사 외에도 화자의 특징을 지시하는 표현들이 어떤 방식으로든 텍스트 전체에 걸쳐 반영되어 있을 것이라고 짐작할 수 있다. 본 연구에서 특히 주목하고 있는 화자명과 화자 지시어 사이의 '거리'가 그 중 하나다. 거리 자체는 언어학적 용법에 속하지 않지만, 거리는 화자 지시어와 화자와의 관련성, 즉 화자 지시어가 화자에 미치는 영향의 정도를 가늠할 수 있는 하나의 요소다. 본 연구는 이처럼, 문법적 용법에 속하지 않으면서도 화자의 원형적 페르소나를 담고 있는 다양한 표현 방식과 패턴들을 찾기 위해 다음과 같은 순서로 분석 작업을 진행하였다.

1. 선행 작업 수행: 별명, 별칭 분류
2. 분석 범위 및 선별 기준 설정
3. 화자 지시어 추출(표본 작업) 및 규칙 분석
4. 규칙 검증(텍스트 확대 분석)
5. 규칙 정리

본 연구에서 화자 지시어를 분류하는 방법은 크게 ‘규칙 기반 방식’과 ‘등장 확률 방식’ 두 가지 방향에서 진행되었다. 전자는 분석 작업에서 수행된 방식으로, 본 연구에서는 샘플링 분석을 통해 기본 패턴 2가지와 예외 패턴 11가지를 우선 분류하고, 규칙 검증 단계를 거쳐 공통 규칙 9가지를 최종 정리하였다. 후자는 규칙별 출현 빈도수를 산출하여 알고리즘이 특정 화자를 정의하기 힘든 상황에서 오차 범위를 최소화하기 위해 시도된 방법이다. 이 작업은 화자 지시어를 모두 추출한 후 결과 및 평가 단계에서 각 규칙별로 화자 지시어 출현 빈도수를 계산, 비교하는 방식으로 진행되었다.

### 1. 선행 작업 - 별명, 별칭 분류

본격적으로 텍스트 분석에 들어가기 전에, 먼저 화자를 직접 지칭하는 별명 혹은 별칭 목록을 작성하였다. 별칭, 별명은 인칭 대명사와 달리 해당 화자를 직접 지시하기 때문에 별도의 분석 작업을 필요로 하지 않으며, 화자명과 매칭된 화자 지시어 목록을 제공하는 것만으로도 기계 학습이 가능하다.

문학 텍스트를 머신러닝에 활용하고자 하는 연구에서 텍스트를 학습에 적용하기 전에 별칭, 별명을 미리 정리해야 하는 이유에는 다음의 사례가 좋은 예시이다. [표 1]에서 볼드체로 표기된 용어는 분석 텍스트에서 두 화자 간에 중복으로 사용된 별칭, 별명들이다. 이 용어들이 지시하는 화자를 구분하지 않은 상태로 텍스트를 학습에 적용하면, 기계가 동일한 별칭, 별명을 사용하는 대화문의 화자를 구별하지 못할 뿐만 아니라 별칭, 별칭을 개별 인물로 판단하여 텍스트 내에 존재하지 않는

가공의 인물을 창출할 수도 있다. 중복 별칭의 경우, 각 별칭마다 해당 화자를 미리 지정해서 학습에 대입하거나, 관련 연구를 활용한 매칭 알고리즘을 개발해서 학습에 적용해야 별칭 중복에 의한 학습 결과 오류를 미연에 방지할 수 있다.

### 2. 분석 범위 및 선별 기준 설정

본 연구는 분석 텍스트 내 등장인물 중에서 사람에게 해당하는 캐릭터로 분석 범위를 제한하였다. 본 연구에서 선정한 텍스트는 판타지에 기반하기 때문에 사람 외에도 동물, 식물과 같이 다양한 유형의 의인화된 캐릭터들이 출현한다. 본 연구에서 인간형 캐릭터를 우선 분석하는 이유는 이후 캐릭터 개발 작업을 효율적으로 진행하기 위함이다. 동물화, 식물화 캐릭터의 경우 다양한 유형의 동식물 모형에 기반하기 때문에 많은 수의 모델 분석이 필요하다. 활용도가 높은 인간형 캐릭터 모델을 먼저 구현한 후 이를 다른 유형의 모델로 확장하는 것이 연구의 효율성을 높일 수 있다.

본 연구에서 분석한 텍스트 내에서 발견된 화자 지시어의 문법적 유형은 인칭 대명사, 별명, 의존명사, 대명사와 명사가 결합된 경우(e.g., they all), 진주어 등이 있었다. 문학 텍스트라는 특성을 고려해서, 분석 초기 단계에 다음과 같이 세 가지 기준을 우선 설정하였다. 각각의 조건들은 텍스트적 지형 내에서 화자 지시어와 그 지시어가 가리키는 대상 사이의 지리적 관계를 나타낸다: 화자명이 화자 지시어로부터 가까운지 아니면 멀리 떨어져 있는지, 화자 지시어가 화자명에 선행하는지 혹은 후행하는지, 화자 지시어가 인용문 안에 포함되어 있는지 아닌지 등이 그것이다.

- 1) 거리: 화자 지시어와 화자명 사이의 거리
- 2) 위치: 화자명이 화자 지시어 앞에 있는지 뒤에 있는지 여부
- 3) 인용문: 화자 지시어가 인용문 안에 들어있는지 아닌지 여부

화자 지시어 중에서도 특히 대명사는 화자명의 반복을 피하기 위한 용도로 주로 사용된다. 이 때문에 화자명에 뒤따르는 후치 지시가 일반적인데, 특이하게도, 문학 텍스트에서는 선행 지시의 사례가 종종 발견되곤 한다. 본 연구에서 채택한 텍스트에서도 47페이지 정도의

표 1. 별칭, 별명 목록  
 Table 1. List of other names and nicknames

Lucy	Lu, <b>Daughter of Eve, queen</b>
Susan	Su, <b>Daughter of Eve, queen</b>
Edmund	Ed, <b>Son of Adam</b> , a little beast, <i>king</i>
Peter	<b>Son of Adam, king</b>
the Professor	an old man
Mrs. Macready	a housekeeper
the Witch	the White Witch, the Queen, the Queen of Narnia, your Majesty, Jadis, the Lady
Father Christmas	a huge man

분량 안에서 총 20개의 선행 화자 지시어가 발견되었다.

### 3. 화자 지시어 추출(표본 작업) 및 규칙 분석

본 연구는 전체 텍스트를 분석하기 전에 1장을 표본 샘플링 범위로 정하고, 화자 지시어가 화자를 지시하는 방식을 분석한 후 기본 규칙 후보 4가지와 예외 규칙 후보 8가지를 우선 분류하였다. 기본 규칙은 화자명과 관계성이 높을 것으로 예상되는 패턴들이다. 예외 규칙에는 화자 지시어들의 문법적 유형과 지시 방식, 지시 조건들을 모두 포함하였고, 화자 지시어가 단수가 아닐 경우, 2인과 3인 이상의 경우로 구분해서 분석하였다.

표 2. 화자 지시어 규칙 1차 분석  
Table 2. Primary analysis of speaker directive rules

기본 규칙	1. 화자명과 화자 지시어의 위치 (선행/후행) 2. 화자명과의 거리가 가까운지 여부 3. 화자 지시어가 사람을 지시하는 경우와 그렇지 않은 경우 4. 화자 지시어가 단수인 경우와 복수인 경우
예외 규칙	1. 지시 대명사 2. 진주어와 가주어 3. 직접 인용문 안에 있는 화자 지시어와 직접 인용문 밖에 있는 화자 지시어 4. 화자 지시어가 화자명 뒤에 있는 경우 5. 화자 지시어가 1인을 지칭하지 않고, 2인 혹은 3인 이상 지칭하는 경우 6. 화자 지시어의 화자명이 이야기 안에 있는 경우와 이야기 밖에 있는 경우(가령, 독자나 저자를 지칭하는 경우) 7. 화자 지시어가 다른 명사와 함께 쓰인 경우 (e.g., They all) 8. 복수형 화자 지시어가 지칭하는 화자들이 텍스트 전체에 흩어져 있는 경우

### 4. 규칙 검증(텍스트 확대 분석)

규칙 검증 작업은 1장에서 추출된 12개의 규칙을 7장까지 확대 분석하는 것으로 진행되었다. 검증 작업은 1장에서 도출된 규칙이 이후의 장에서도 지속해서 출현하는지, 그리고 1장에는 없었던 새로운 규칙이 존재하는지를 확인하기 위함이다. 만약 1장에서만 나타나고 이후의 장에서 출현하지 않는 규칙이라면 일반 규칙에서 제외되어야 마땅하고, 1장에는 없었다 해도 이후의 장에서 새로운 규칙이 나타나고, 또 그 규칙이 지속해서 출현한다면 최종 규칙에 포함될 필요가 있다. 추가 분석 결과 새로운 유형의 규칙은 나타나지 않았다.

### 5. 규칙 정리

아래는 텍스트 샘플링 및 추가 텍스트 분석 작업을 거친 후 재정리된 화자 지시어 분석을 위한 9가지 규칙이다.

- 1) 위치
- 2) 거리
- 3) 대명사
- 4) 가주어/진주어
- 5) 인용문
- 6) 화자수
- 7) 등장인물 외 지시
- 8) 복합 단어 지시
- 9) 화자명 분산

첫 번째 규칙인 ‘위치’는 화자 지시어가 놓인 위치를 의미한다. 분류작업에서는 대부분의 화자 지시어가 화자명에 뒤따라 나오기 때문에 화자 지시어가 화자명 앞에 놓인 경우만 “선행”으로 표기하였다. 두 번째 규칙, ‘거리’는 화자명과 화자 지시어 사이에 놓인 공백(blank)의 개수를 기준한다. 화자명과 화자 지시어 사이의 거리는 화자 지시어가 화자명에 미치는 영향을 가늠하기 위한 척도로 활용될 수 있으며, 결과적으로 화자명과 가장 가까운 화자 지시어를 선별하여 캐릭터 분석 작업의 효율을 가져오기 위함이다. 공백수를 산출하면 화자명과 화자 지시어 사이에 놓인 단어의 개수도 함께 파악된다. 세 번째 규칙, ‘대명사’는 가장 빈번하게 나타나는 화자 지시어의 문법적 유형이다. 본 연구의 목적이 문학 텍스트 내의 등장인물을 가상 캐릭터로 활용하는 것이기 때문에, 대명사 중에서도 지시 대명사나 지시어는 제외하고 인칭 대명사만 분석에 포함하였다. 네 번째 규칙인 ‘가주어/진주어’는 화자 지시어가 해당 화자의 주격 대명사로 표현되지 않고 공통적으로 가주어 ‘It’으로 표기된 경우다. 가주어 ‘It’은 이전 연구에서도 기계가 화자를 분별하지 못해서 캐릭터 분석에서 누락된 사례가 있다. 본 연구에서는 가주어 ‘It’ 자체는 의미를 가지지 않기 때문에 분석에서 제외하고 ‘for’ 혹은 ‘that’의 형식으로 뒤따르는 진주어만 분석하였다. 다섯 번째 규칙, ‘인용문’은 문학 텍스트에서 화자의 대화 내용이 인용문 안에 포함되지 않고 서술되는 사례가 종종 나타남에 따라 분류 기준으로 선정되었다. ‘인용문’은

동일한 유형의 화자 지시어라 해도 화자 지시어가 인용문 안에 포함되어 있는지 아닌지의 여부에 따라 전혀 다른 화자를 지시할 수 있기 때문에 분석에서 중요한 위치를 차지한다. 본 연구에서는 직접 인용문 ‘안’에 있는 화자 지시어와 직접 인용문 ‘밖’에 있는 화자 지시어를 구분하고, ‘직접 인용문 안 또는 밖’이라는 분류 카테고리 추가하여 분석하였다. ([표-3] 참조)

여섯 번째 규칙인 ‘화자수’에는 화자 지시어가 지시하는 화자의 수가 복수인 경우를 포함한다. 본 연구는 화자 지시어가 2인을 초과해서 지칭할 경우 캐릭터와의 연관성이 낮아진다고 보고, 2인까지의 화자 지시어 수를 산출하기 위해 1인을 지칭하는 경우를 포함해서, 2인 사이의 대화와 3인 이상의 대화로 각각 구분, 분석을 수행하였다. 일곱 번째 규칙, ‘등장인물 외 지시’는 화자 지시어가 지시하는 화자명이 이야기 안에 있는 경우와 이야기 밖에 있는 경우를 구분하기 위함이다. 여덟 번째 규칙인 ‘복합 단어 지시’는 화자 지시어가 2개 이상의 복합 단어로 구성된 경우다. 본 연구의 분석 텍스트 내에서 복합 단어로 표현된 화자 지시어는 “they all”과 “you all” 두 종류뿐이었지만, 두 지시어 모두 동일한 대상을 지칭한다는 점에서, 복합 단어 지시의 중요도가 결코 낮다고 볼 수 없어서 규칙에 포함하였다. 마지막으로, 아홉 번째 규칙인 ‘화자명 분산’은 화자 지시어가 지시하는 화자명들이 텍스트 전반에 흩어져 있는 경우다.

표 3. 화자 지시어가 직접 인용문 안에 있는 경우와 직접 인용문 밖에 있는 경우의 한 예

Table 3. An example of a case where a speaker directive is within a direct quotation and outside a direct quotation

직접 인용문 밖에 있는 경우	“I hope you know what I mean by a voice sounding pale.” (10장, 113쪽) → 이 문장에서 I는 narrator를 지시.
직접 인용문 안에 있는 경우	“And then something made Peter say, “. . . I was angry with him and I think that helped him to go wrong.”” (12장, 141쪽) → 이 문장에서 I는 Peter 본인을 지시.

#### IV. 결과 및 평가

평가는 규칙별 출현 빈도수에 따라 규칙의 중요도를 가늠하는 방향으로 진행되었다. 이 중 규칙 2 ‘거리’는 빈도수와 무관하기 때문에 해당 평가 기준을 적용하지

않고 거리 데이터의 활용 방안을 중심으로 평가되었다.

##### 1. 규칙별 출현 빈도수 산출

만약 어떤 규칙이 텍스트 전체에서 단지 한 번만 출현한다면 이 규칙은 공통 규칙으로 보기 어렵다. 따라서, 샘플링 과정에서 선별된 규칙들을 공통 규칙으로 적용할 수 있는지를 판별하기 위한 작업이 필요하다.

이를 위해, 본 연구는 전체 화자 지시어를 대상으로 규칙별 출현 빈도수를 산출하는 작업을 진행하였다. 출현 빈도수는 샘플링 과정에서 선별된 규칙들을 톨킨의 『반지의 제왕』 1권의 1장에 대입하여 9가지 규칙들의 유효성을 확인한 후, 나니아 연대기 1권에서 17장에 적용하는 순으로 검증 작업을 진행하였다. [표-4]는 추출된 화자 지시어의 규칙별 출현 빈도수 및 화자 지시어 총합을 나타낸다.

출현 빈도수 산출 작업은 선별된 화자명과 화자 지시어 사이의 관계가 이미 정의된 규칙 안에 포함되는지, 그리고 선별된 규칙들이 공통 규칙으로 정의할 만큼 텍스트에서 사용되었는지를 확인하기 위함이다. 가령, [표-4]에서 규칙 4 가주어/진주어는 『나니아 연대기』 제1권 17장 안에서 총 2회 출현, 전체 화자 지시어 수를 기준으로 백분율로 환산했을 때 0.16%로, 규칙에서 제외해도 무방할 정도의 확률을 보였다. 전체 화자 지시어 대비 10% 미만의 확률을 보이는 규칙 1, 7, 8, 9 또한 규칙에서 제외될 수 있다.

표 4. 규칙별 출현 빈도수

Table 4. Frequency of appearance by rules

No.	규칙명	출현 빈도수	통계
규칙 1	위치(선행)	20회	1.57%
규칙 2	거리	해당 없음	-
규칙 3	대명사	1274회	100%
규칙 4	가주어/진주어	2회	0.16%
규칙 5	인용문	직접인용문: 671회	52.67%
규칙 6	화자수	1인 지시 (1014회)	79.59%
		2인 지시 (57회)	4.47%
		3인 이상 (205회)	16.09%
규칙 7	등장인물 외 지시	34회	2.67%
규칙 8	복합 단어	12	0.94%
규칙 9	화자명 분산	126	9.89%
화자 지시어 총합		1274개	

## 2. 출현 빈도수에 따른 각 규칙별 평가

각 규칙별 출현 빈도수는 빈도수 산출과 무관한 ‘거리’를 제외하면, 각각 대명사 > 화자수(1인 지시) > 인용문 > 화자수(3인 이상) > 화자명 분산 > 화자수(2인 지시) > 등장인물 외 지시 > 위치(선행) > 복합 단어 지시 > 가주어/진주어 순으로 나타났다.

### 1) 위치

본 텍스트 분석에서는 1권의 1~7장에서만 총 1,274개의 화자 지시어 중에 20개가 선행하였다. 이 중 1장에서 17개, 4장과 7장에서 각각 2개, 1개의 선행하는 화자 지시어가 출현하였다. 이는 이후의 장에서도 선행하는 화자 지시어가 그다지 많이 출현하지 않을 수 있음을 가정하게 한다. 1장에서 이처럼 상대적으로 많은 수의 선행 화자 지시어가 출현한 것은 작가의 글쓰기 스타일 혹은 1장이 서론의 역할을 대신한다는 점에서 그 원인을 찾을 수도 있을 것이다. 이 부분은 본 논문 외의 영역이므로 언급을 생략한다.

### 2) 거리

문학 텍스트를 활용하고자 하는 머신러닝 학습에서 ‘거리 데이터’의 이점은 분석해야 할 화자 지시어가 많을 때 화자와의 관계성이 높은 것들만 추려서 학습에 사용할 수 있다는 점이다. 앞서, 화자 지시어가 화자명과 같은 문장 안에 있을 때는 화자명과 동급으로 취급할 수 있다고 언급한 바와 같이, 화자명과 화자 지시어 사이의 거리 수치가 낮으면 낮을수록 화자 지시어가 화자의 페르소나에 미치는 영향 혹은 화자와의 관계성이 높아진다고 볼 수 있다. 만약 화자명과 화자 지시어 사이의 거리값이 100 이상이면서 그 사이에 다른 화자 지시어가 포함되어 있다면, 거리값이 100인 화자 지시어는 사실상 화자에 거의 영향을 미치지 않는다고 보고 학습 데이터에서 제외해도 무방하다.

본 연구에서 산출한 거리값은 문학 텍스트 내의 캐릭터를 가상현실이나 메타버스 환경에 구축하기 위해 특성을 부여하는 작업에서 각각의 화자 지시어에 가중치를 부여하는 방식으로 활용될 수 있다. 이 경우 각 용어에 동일한 가중치를 부여하는 것보다 더 높은 확률로 문학 텍스트에서 표현하고자 하는 등장인물의 페르소나에 근접할 수 있다.

### 3) 대명사

서론에서 언급한 바와 같이, 화자를 지시하는 지시어에는 대명사 외에도 별명, 별칭, 동격명사, 명사구, 서술명사 등이 있지만, 본 연구에서 동사 관련 용어 분석을 제외하고, 별명, 별칭을 따로 목록화하면서, 결과적으로 분석된 모든 화자 지시어가 대명사에 해당하는 결과로 나타났다.

### 4) 가주어/진주어

가주어/진주어는 총 1274개 화자 지시어 중 1장에서만 2회 출현, 0.16%의 빈도수를 보였다. 이는 저자가 인간형 화자를 지시하는 데는 가주어/진주어 형식을 거의 사용하지 않았음을 의미한다. 이에 따라, 규칙 4는 규칙 적용 여부를 재고해야 할 필요가 있는 것으로 평가되었다.

### 5) 인용문

인용문 중 직접 인용문은 671회, 약 52.67%의 확률을 보였다. 이는 직접 인용문과 간접 인용문에서 화자 지시어가 서로 비등하게 사용되었음을 의미한다. 챗봇을 활용한 이전 연구에서 본 연구자는 챗봇이 인용부호 이외의 문장을 대화문으로 인식하지 않아서 학습에서 제외된 바 있다. 하지만, 위 결과는 기계 학습을 통해 문학 캐릭터를 산출할 때 간접 인용문의 학습이 직접 인용문 만큼이나 중요하다는 것을 시사한다. 이 결과는 기계 학습이 제대로 수행되기 위해서는 문학 텍스트를 학습에 도입하기 이전에, 간접 인용문을 직접 인용문의 형식으로 바꾸거나, 전체 학습 텍스트에서 인용부호를 삭제한 후 모든 텍스트를 학습하게 하는 등의 전처리 작업의 필요성을 제기한다.

### 6) 화자수

[표 4]에서, 화자 지시어가 지칭하는 화자수 분석 결과가 1인을 지칭하는 화자 지시어는 대략 80%, 2인 이상 지시하는 화자 지시어는 약 20% 산출된 것을 확인할 수 있다. 데이터가 방대하면 방대할수록 학습에 필요한 시스템의 사양이 높아지고 소요되는 시간이 늘어난다는 점을 감안할 때, 이 결과는 20%에 해당하는 2인 이상의 화자 지시어를 학습 자체에서 제외하는 방향보다는, 1인을 지칭하는 화자 지시어를 먼저 학습하게 한 후, 2인 이상의 화자 지시어를 추가 학습하게 하는 방식으로,



학습 기간을 조정하는 데 활용할 수 있다. 우선순위는 각 규칙들간의 가중치 적용으로 조율 가능하다.

#### 7) 등장인물 외 지시

본 분석 범위에서는 'I' 혹은 'You', 'We' 와 같은 대명사가 등장인물이 아닌 이야기 밖의 독자나 저자를 지시하는 화자 지시어로 등장하였다. 본 연구는 분석 작업을 통해서, 분석 텍스트 중 특히 7장에서 '등장인물 외 지시' 화자어가 7장에서 산출된 총 화자 지시어 수 108개 중에서 16개, 즉 14.81%의 확률로 출현하는 것을 확인할 수 있었다. 이는 지금까지 분석한 장 중에서도 가장 높은 확률을 보이는 횟수이며, 1장에서 4개의 '등장인물 외 지시'어가 나온 것을 감안할 때, 이런 결과는 이후의 장 분석에서 '등장인물 외 지시'어가 더 많은 확률로 나타날 수도 있다는 것을 가늠케 한다. 또한, 이런 결과는 단지 저자가 서론 격인 1장에서만 저자나 독자를 언급하기 위해 '등장인물 외 지시'어를 사용한 것이 아니었음을 시사한다. 분석과 무관하게, 이야기 밖에 있는 독자나 저자를 기계 학습에 도입해서 캐릭터 분석을 수행할 필요성이 있는지의 여부는 가상 공간이나 메타버스 환경에서 문학 캐릭터를 어떻게 활용할지에 달린 것으로 보아야 할 것이다.

#### 8) 복합 단어 지시

1장에서 나타난 "they all"에 의해 규칙 8은 복합 단어로써 화자를 지시하는 사례로 규칙 후보로 선정되었다. 본 분석에서 화자명 분산 사례는 'they all' 과 'you all' 2종류였으며, 'they all'은 1~7장에서 총 11회, 'you all'은 총 1회 출현하였다. 전자는 간접 인용문에서, 후자는 직접 인용문에서의 표기 차이로, 사실상 두 경우 모두 동일한 화자들을 지시한다. 이 두 단어는 작가의 글쓰기 특성상, 해당 장에서 지시하고자 하는 화자가 모두 언급되지 않았을 때 사용되었다. ('they all', 'you all' 모두 네 명의 화자, 즉 Lucy, Susan, Edmund, Peter 를 지칭하는데, 해당 장에서 특정 인물이 언급되지 않았을 경우, 저자가 they 혹은 you 뒤에 all을 붙여서 네 명 모두를 지시한다는 것을 분별하기 위해 사용함)

이처럼 저자의 글쓰기 스타일을 미리 파악한다면, 위 두 단어도 별칭, 별명 목록처럼 화자를 직접 지시해서 기계 학습을 수행할 수 있다. 하지만, 본 텍스트 분석 범위에서는 출현하지 않았지만, 'both of them'이나 'a

part/parts of them', 'one of them'과 같이, 여러 화자 중에 일부만을 지시하는 복합 단어 지시어들은 문맥 화인을 통해서만 화자를 추적할 수 있다. 이런 사례가 발견될 경우, 출현 빈도와 중요도에 따라 별칭, 별명의 경우처럼 별도의 수작업이 병행될 필요가 있다.

#### 9) 화자명 분산

텍스트 전체에 화자 지시어가 지칭하는 화자명이 분산된 경우는 총 1274개의 화자 지시어 중에 126개로 나타났다. 이는 총 화자 지시어에서 9.89%의 확률이며, 규칙 후보 중에서 빈도수와 무관한 거리를 제외하고, 화자수를 하나의 규칙으로 간주하면, 4번째로 빈도수가 높은 규칙에 해당한다. '화자명 분산' 빈도수 약 10%는 전체 화자 지시어 수에서 무시할 수 없는 수치로 간주되며, 학습의 난이도와 화자 지시어의 중요도에 따라 규칙 적용 여부를 가늠할 수 있을 것이다.

### 3. 정리

화자 지시어가 화자명과 가장 밀접한 관계를 가지는 조건은 '화자명에 뒤따르면서 가장 가까운 곳에 위치한 경우'다. 화자명과 화자 지시어가 같은 문장 안에 위치하면 화자 지시어는 화자명과 거의 동급으로 볼 수 있다.

규칙은 내용을 알지 못해도 항상 같은 결과가 산출되어야 규칙이라고 말할 수 있다. 학습 알고리즘이 내용을 모르고도 주어진 규칙만으로 화자 지시어가 지칭하는 화자를 찾을 수 있게 하기 위해서는 학습에 들어가기 전에 등장인물의 목록(화자명) 및 각 인물의 발화문장(utterance sentence)이 함께 매칭된 데이터를 기본 자료로 제공해 주어야 한다. 또, 등장인물의 등장 확률을 개별 산출해서 학습 자료로 제공해 주는 것도 학습의 효율성을 높일 수 있는 좋은 방법이다. '등장 확률 분석'은 분석 텍스트 범위 내에서 각각의 인물들의 등장 횟수와 등장인물 간의 거리를 포함한다. 이 자료들은 학습 알고리즘이 특정 등장인물로 압축하기 어려운 화자 지시어를 맞닥뜨렸을 때 모호한 값에 대한 선택 범위를 줄여서 학습을 용이하게 하고 정확도를 높이는 기준 데이터로 작용할 수 있다. 알고리즘이 90%의 수준에서 화자 지시어에 결합하는 화자를 찾아낸다면 학습에 성공한 것으로 간주한다.

## V. 결론 및 향후 연구

기계 학습을 위해서는 명확한 규칙의 제시가 필요하다. 모호함은 기계 학습에 혼란을 가져와서 유의미한 결과를 도출할 수 없게 한다. 본 연구에서 제시한 규칙을 활용하면 문학 텍스트를 머신 러닝에 활용하는 데 시행착오를 거치지 않고 질적으로 우수한 결과를 산출할 수 있을 것으로 기대한다.

본 연구는 처음, 대명사 이외에도 화자 지시어가 있을 것으로 가정하고 출발하였기 때문에, 본 연구에서 정리한 화자 지시어 규칙 목록에 대명사가 포함되었다. 분석 결과, 별명, 별칭을 제외한 화자 지시어가 모두 대명사인 것으로 나타났으며, 이에 따라, 본 연구 주제를 화자 지시어가 아니라 대명사 연구로 변경하고, 규칙에서 대명사를 제외해야 할 필요성이 연구 과정 중에 제기되기도 하였다. 하지만, 본 연구자는 화자 지시어라는 용어가 대명사를 포함하면서도 더 넓은 범위의 지시어들을 포괄할 수 있으며, 본 연구가 이후 형용사 및 부사, 동사로 그 영역을 확대해 나아가야 진정한 문학 캐릭터의 가상화가 가능하다고 판단하고, 본 연구의 주제를 ‘화자 지시어’로 유지하는 것으로 최종적으로 결정하였다.

본 실험을 통해서 발견된 유의미한 요소 중의 하나는 문학 작품에서 화자 지시어의 규칙들이 1장 안에 모두 들어있었다는 점이다. 지금까지 이와 관련된 연구—1장이 일련의 서론으로서 역할 한다는 것에 관한—가 부재했다는 것은 문학 텍스트에 관한 연구가 구조적 접근보다는 내용적인 측면을 중심으로 연구되어왔다는 것을 시사한다. 본인은 본 연구의 이와 같은 방향이 문학 텍스트 연구에 새로운 방향을 제시해 줄 수 있으며, 내용 중심의 연구에서는 포착할 수 없었던 문학 텍스트의 잠재적 가치를 발굴할 수 있는 좋은 계기로 작용할 것으로 전망한다.

본 연구는 문학 텍스트의 등장인물을 머신러닝 학습에 기반한 가상 캐릭터 구현에 활용하기 위한 목적으로 수행되었다. 문학 텍스트에 표현된 등장인물의 페르소나가 반영된 가상 캐릭터를 구현하기 위해서는 해당 텍스트 내에서 특정 캐릭터가 한 ‘말’과 ‘행동’이 드러나는 문장을 분석해야 한다. 하지만, 이보다 더욱 중요한 것은 ‘기계가 이해할 수 있는 방식으로 학습 텍스트를 제 공해야 한다’는 점이다. 지난 연구에서 기계가 대명사가

지시하는 화자를 스스로 판별하지 못하여 수동으로 분석 작업을 수행한 것이 본 연구를 촉발하는 계기가 되었다. 본 연구는 이를 해결하기 위한 후속 연구로서, 향후 기계 학습 연구에서는 기계가 스스로 화자 지시어가 지시하는 화자를 직접 판별할 수 있도록 9가지 화자 지시어 규칙을 직접 적용하여 그 결과를 확인하고자 한다.

본 논문에 분석 데이터로 포함하지는 않았지만, 본 연구자는 작가들 사이에서 화자 지시어 이용에 차이가 있는지를 확인하기 위해 톨킨의 “반지의 제왕” 일부를 분석한 바 있다. 본 연구에서 추출한 9가지 규칙을 톨킨의 텍스트에 적용했을 때 규칙의 차이는 없었지만, “화자 지시어”의 쓰임새가 두 작가에게서 조금은 다르게 나타나는 것을 확인할 수 있었다. 분석 작가를 확대하면, 아마도 대다수의 다른 작가들에게서도 이와 유사한 결과를 얻을 것으로 보인다. 기존 분석에서 드러나지 않은 화자 지시어 용법이 나타날 경우, 이 용법은 더 많은 텍스트 분석을 거친 후에 사용 빈도가 높으면 공통 규칙으로 추가되고, 특정 작가에게서만 나타나면 분석에서 제외될 수도 있을 것이다. 중요한 점은, 본 연구에서 분석한 루이스의 텍스트와 톨킨의 일부 텍스트에서 추가할 만한 규칙을 아직 발견하지 못했다는 점이다. 이것은 본 연구에서 제시하는 ‘9가지 화자 지시어 규칙’이 일정 정도는 문학 텍스트에 공통 규칙으로 적용될 수 있다는 가능성을 내포한다. 다른 작가, 다른 문학 텍스트로의 확대 분석 및 검증 작업은 향후 연구 과제가 될 것이다.

## References

- [1] Hyeongu Jeon, Kichul Jung, Kyoungah Kwon, and Insung Lee. “Machine Learning Language Model Implementation Using Literary Texts.” *The Journal of the Convergence on Culture Technology (JCCT)* 7.2 (2021): 427-436.
- [2] C. S Lewis. (2001). *The Chronicles of Narnia*. New York: HarperCollins Publishers.
- [3] Kang, Seung-Shik, Yun, Bo-Hyun, and Woo, Chong-Woo. “Antecedent Decision Rules of Personal Pronouns for Coreference Resolution.” *The KIPS Transactions : Part B. The Korea Information Processing Society*, 11.2 (2004): 227-232.
- [4] Carbonell, Jaime G., and Ralf D. Brown. “Anaphora

- resolution: a multi-strategy approach.” Coling Budapest 1988 Volume 1: International Conference on Computational Linguistics. 1988.
- [5] Carter, David. Interpreting anaphors in natural language texts. Halsted Press, 1987.
- [6] Hobbs, Jerry R. “Pronoun resolution.” ACM SIGART Bulletin 61 (1977): 28-28.
- [7] Baldwin, Breck. “CogNIAC: high precision coreference with limited knowledge and linguistic resources.” Operational factors in practical, robust anaphora resolution for unrestricted texts. 1997.
- [8] Cheoneum Park, Changki Lee. “Coreference Resolution for Korean Pronouns using Pointer Networks.” Journal of KIISE, 44.5 (2017): 496-502.
- [9] C. Park, G. H. Choi, and C. Lee, “Korean Coreference Resolution with Guided Mention Pair Model using the Deep Learning,” Proc. Of the KIISE Korea Computer Congress 2015, pp. 693-695, 2015. (in Korean)