



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

文學碩士 學位論文

대화코퍼스를 활용한 셰익스피어 사랑 비극의  
감정 분석과 네트워크 분석

Sentiment Analysis and Network Analysis of Dialogue  
Corpora in Shakespeare's *Love Tragedies*



指導教授 張 世 銀

2018年 8月

韓國海洋大學交 大學院

英語 英文學科

李 京 恩



본 논문을 이경은의 문학석사 학위논문으로 인준함.



위원장

홍 옥 숙



위원

송 원 문



위원

장 세 은



2018년 06월05일

한국해양대학교 대학원



# 목 차

List of Tables .....	iii
List of Figures .....	vi
Abstract .....	viii
<b>1. 서론 .....</b>	<b>1</b>
1.1. 연구 목적과 배경 .....	1
1.2. 연구 질문 .....	4
<b>2. 선행 연구 .....</b>	<b>6</b>
2.1. 코퍼스언어학과 세익스피어 극 .....	6
2.2. 감정 분석 .....	9
2.3. 네트워크 분석 .....	14
2.4. 감정 분석 사전 .....	16
2.4.1. AFINN .....	16
2.4.2. VADER .....	17
<b>3. 데이터와 방법론 .....</b>	<b>21</b>
3.1. 데이터 .....	21
3.1.1. 사랑 비극 .....	21
3.1.1.1. Romeo and Juliet .....	24
3.1.1.2. Othello, the Moor of Venice .....	26
3.1.1.3. Antony and Cleopatra .....	28
3.1.2. 사랑 희극 .....	29
3.1.2.1. The Two Gentlemen of Verona .....	31
3.1.2.2. A Midsummer Night's Dream .....	33
3.1.2.3. As You Like It .....	35

3.2. 방법론 .....	38
3.2.1. 감정 분석 .....	38
3.2.2. 네트워크 분석 .....	40
<b>4. 사랑 비극의 감정 분석 .....</b>	<b>42</b>
4.1. 사랑 비극 간의 비교 .....	43
4.2. 사랑 희극 간의 비교 .....	59
4.3. 사랑 비극과 사랑 희극의 비교 .....	60
4.4. 사랑 비극의 주요 등장인물 간 비교 .....	62
4.4.1. 남성 주인공과 여성 주인공의 감정 분석 .....	63
4.4.1.1. Romeo and Juliet .....	63
4.4.1.2. Othello and Desdemona .....	69
4.4.1.3. Antony and Cleopatra .....	75
4.4.2. 막 별 주요 등장인물 간 분석 .....	81
4.4.2.1. Romeo and Juliet .....	81
4.4.2.2. Othello, the Moor of Venice .....	84
4.4.2.3. Antony and Cleopatra .....	86
4.5. 요약 .....	88
<b>5. 사랑 비극의 네트워크 분석 .....</b>	<b>91</b>
5.1. 사랑 비극 내 주요 등장인물 간 네트워크 분석 .....	91
5.1.1. Romeo and Juliet .....	91
5.1.2. Othello, the Moor of Venice .....	93
5.1.3. Antony and Cleopatra .....	96
5.2. 시간시각화와 관계시각화의 비교 .....	97
5.3. 요약 .....	98
<b>6. 결론 .....</b>	<b>99</b>
6.1. 요약 .....	99
6.2. 향후 과제 .....	101
<b>참고 문헌 .....</b>	<b>102</b>

## List of Tables

Table 1	General Information of Three Love Tragedies .....	23
Table 2	Dialogue Corpus Information of <i>Romeo and Juliet</i> .....	24
Table 3	General Information of Characters in <i>Romeo and Juliet</i> .....	25
Table 4	Dialogue Corpus Information of <i>Othello, the Moor of Venice</i> .....	26
Table 5	General Information of Characters in <i>Othello, the Moor of Venice</i> .....	27
Table 6	Dialogue Corpus Information of <i>Antony and Cleopatra</i> .....	28
Table 7	General Information of Characters in <i>Antony and Cleopatra</i> .....	29
Table 8	General Information of Three Love Comedies .....	30
Table 9	Dialogue Corpus Information of <i>The Two Gentlemen of Verona</i> .....	31
Table 10	General Information of Characters in <i>The Two Gentlemen of Verona</i> .....	32
Table 11	Dialogue Corpus Information of <i>A Midsummer Night's Dream</i> .....	33
Table 12	General Information of Characters in <i>A Midsummer Night's Dream</i> .....	34
Table 13	Dialogue Corpus Information of <i>As You Like It</i> .....	35
Table 14	General Information of Characters in <i>As You Like It</i> .....	37
Table 15(a)	Normalization example of Three Love Tragedies .....	38
Table 15(b)	Normalization example of Three Love Tragedies .....	39
Table 16	Cumulative VADER Valences of <i>Romeo and Juliet</i> .....	44

<b>Table 17</b>	Cumulative VADER Valences of <i>Othello, the Moor of Venice</i> .....	45
<b>Table 18</b>	Cumulative VADER Valences of <i>Antony and Cleopatra</i> .....	46
<b>Table 19</b>	Dispersion of Dialogue including ‘Love’ in <i>Romeo and Juliet</i> .....	47
<b>Table 20</b>	Dialogue including ‘Love’ in <i>Romeo and Juliet</i> .....	48
<b>Table 21</b>	Dispersion of Dialogue including ‘Love’ in <i>Othello, the Moor of Venice</i> .....	49
<b>Table 22</b>	Dialogue including ‘Love’ in <i>Othello, the Moor of Venice</i> .....	49
<b>Table 23</b>	Dispersion of Dialogue including ‘Love’ in <i>Antony and Cleopatra</i> .....	51
<b>Table 24</b>	Dialogue including ‘Love’ in <i>Antony and Cleopatra</i> .....	51
<b>Table 25</b>	Rank of Cumulative VADER Valences of <i>Romeo and Juliet</i> .....	52
<b>Table 26</b>	Rank of Cumulative VADER Valences of <i>Othello, the Moor of Venice</i> .....	54
<b>Table 27</b>	Rank of Cumulative VADER Valences of <i>Antony and Cleopatra</i> .....	57
<b>Table 28</b>	Normalization for VADER and AFINN Valences .....	64
<b>Table 29(a)</b>	Comparing Valences of VADER and AFINN of Romeo-Juliet .....	65
<b>Table 29(b)</b>	Comparing Valences of VADER and AFINN of Juliet-Romeo .....	65
<b>Table 30(a)</b>	Comparing Valences of VADER and AFINN of Romeo-Juliet .....	66
<b>Table 30(b)</b>	Comparing Valences of VADER and AFINN of Juliet-Romeo .....	66
<b>Table 31</b>	VADER and AFINN Valences and Cumulative VADER and AFINN Valences of Romeo-Juliet .....	67
<b>Table 32(a)</b>	Comparing Valences of VADER and AFINN of Othello-Desdemona	

.....	71
<b>Table 32(b)</b> Comparing Valences of VADER and AFINN of Desdemona–Othello .....	71
<b>Table 33(a)</b> Comparing Valences of VADER and AFINN of Othello–Desdemona .....	72
<b>Table 33(b)</b> Comparing Valences of VADER and AFINN of Desdemona–Othello .....	72
<b>Table 34</b> VADER and AFINN Valences and Cumulative VADER and AFINN Valences of Othello–Desdemona .....	73
<b>Table 35(a)</b> Comparing Valences of VADER and AFINN of Antony–Cleopatra .....	77
<b>Table 35(b)</b> Comparing Valences of VADER and AFINN of Cleopatra–Antony .....	77
<b>Table 36(a)</b> Comparing Valences of VADER and AFINN of Antony–Cleopatra .....	78
<b>Table 36(b)</b> Comparing Valences of VADER and AFINN of Cleopatra–Antony .....	78
<b>Table 37</b> VADER and AFINN Valences and Cumulative VADER and AFINN Valences of Antony–Cleopatra .....	79
<b>Table 38</b> Comparing between Time visualization and Network visualization .....	97

## List of Figures

Fig. 1 Love related lexical items with occur in the comedies and tragedies .....	7
Fig. 2 Average Word Valence : Tragedies vs Comedies .....	10
Fig. 3 Hamlet's Sentiment Valence Sum .....	11
Fig. 4 Comparing between Sentiment Valence of Hamlet-Gertrude .....	12
Fig. 5 The Social Networks for <i>The Taming of the Shrew</i> and <i>The Tragedy of Macbeth</i> .....	15
Fig. 6 Network Structure of Three Love Tragedies .....	16
Fig. 7 AFINN Sentiment Dictionary .....	17
Fig. 8 VADER Sentiment Dictionary .....	18
Fig. 9 Sentiment Score form VADER and 11 other sentiment analysis tools .....	18
Fig. 10 VADER performance compared to human raters .....	19
Fig. 11 Network Platform of NetMiner 4.0 .....	40
Fig. 12 Sentiment Analysis of Three Love Tragedies .....	43
Fig. 13 Sentiment Analysis of Three Love Comedies .....	59
Fig. 14 Sentiment Analysis of and Three Love Tragedies and Three Love Comedies .....	61
Fig. 15 Comparing Cumulative Sentiment Valences using AFINN and VADER of Romeo-Juliet .....	64
Fig. 16 Comparing Cumulative Sentiment Valences using AFINN and VADER of Othello-Desdemona .....	70
Fig. 17 Comparing Cumulative Sentiment Valences using AFINN and VADER of Antony-Cleopatra .....	76
Fig. 18 Comparison between All Major Characters in <i>Romeo and Juliet</i> .....	81
Fig. 19 Comparison between All Major Characters in <i>Othello, the Moor of Venice</i> .....	84

Fig. 20 Comparison between All Major Characters in *Antony and Cleopatra* .....86

Fig. 21 Network Analysis of Six Major Characters in *Romeo and Juliet* .....91

Fig. 22 Network Analysis of Six Major Characters in *Othello, the Moor of Venice* .....93

Fig. 23 Network Analysis of Four Major Characters in *Antony and Cleopatra* .....96



# Sentiment Analysis and Network Analysis of Dialogue Corpora in Shakespeare's Love Tragedies

Lee, Kyung Eun

Department of English and Literature  
Graduate School of Korea Maritime and Ocean University

## Abstract

This study explores a new sentiment analysis of dialogue corpora in Shakespeare's three love tragedies (*Othello*, *the Moor of Venice*, *Anthony and Cleopatra* and *Romeo and Juliet*). It also investigates sentiment valence of male and female main characters by comparing Valence Aware Dictionary for Sentiment Reasoning (VADER) and AFINN sentiment graphs. Furthermore, it shows some characteristics of major characters through visualizing their sentiment graphs and networks in the three love tragedy plays. Some characteristics of major characters, which cannot be found in sentiment analysis, can be identified by network relationships among major characters visualizing their intensity or strength.

Knowing who is speaking to whom allows the flow of sentiment relationships to be tracked between characters and within plays with clear time-lines. In previous research, Nalisnick (2013) used the AFINN word list

for sentiment analysis. My approach employs VADER suggested by Hutto and Gilbert (2014) to calculate the sentiment valences of dialogue corpora. VADER is a model used for text sentiment analysis that represents both positive or negative polarity and intensity (strength) of characters' emotion. In order to visualize the network structures, NetMiner 4.3 tools were used to create sentiment networks connecting each of the major characters.

Here are my four research questions to study my aims:

1. Are there prominent sentiment patterns between three love-tragedy plays and three love-comedy plays?
2. Are there any clues to distinguish tragedy from comedy in the sentiment analysis?
3. Which sentiment dictionary is more adequate for sentiment analysis?
4. How do we find the sentiment relationships between major characters?

This study will contribute to clearly distinguishing and analyzing human relationships described in the text written in English. Moreover, it can be extended to applying this approach to the sentiment relationships in plays, novels, or other registers written in other languages.

**KEY WORDS:** Shakespeare 셰익스피어; Love Tragedy 사랑 비극; Dialogue Corpus 대화코퍼스; Sentiment Analysis 감정 분석; Network Analysis 네트워크 분석.

## 초 록

본 연구는 셰익스피어의 비극 작품들 중 세 편의 사랑 비극 (*Romeo and Juliet*, *Othello*, *the Moor of Venice*, *Antony and Cleopatra*)에서 모든 등장인물들의 대화코퍼스를 구축하여 Hutto and Gilbert (2014)가 개발한 Valence Aware Dictionary for Sentiment Reasoning (VADER) 감정 분석 사전과 Nalisnick (2013)이 연구에 사용한 기존의 AFINN 감정 분석 사전의 비교를 통해 비극 작품 및 등장인물 간의 대화코퍼스를 감정 분석하고 NetMiner 4.0을 통해 주요 등장인물 간의 네트워크 분석을 하여 주요 등장인물들이 가지는 네트워크가 부정적인지 긍정적인지를 살펴본다. 이러한 비극 작품 및 등장인물 간의 특징을 시간의 흐름에 따른 시간시각화로 살펴보고 이러한 시간시각화에서는 발견할 수 없는 등장인물 간 연결 강도를 네트워크 분석에 따른 관계시각화로 살펴본다.

본 연구를 통해 다음의 네 가지 연구 질문의 답을 찾으려고 하였다. 본 연구는 첫째, 사랑을 주제로 한 비극 대화코퍼스의 감정 분석을 통해 비극의 특징을 보여주는 감정 패턴을 발견 할 수 있는가? 선그래프의 시각화를 통하여 관찰할 수 있는 패턴이 몇 가지가 있는지를 고려한다. 둘째, 감정 분석을 통해 비극과 희극의 차이점을 드러낼 수 있는 패턴 또는 단서가 있는가? 관찰된 사랑 비극의 감정 패턴이 사랑을 주제로 한 희극 세 편 (*A Midsummer Night's Dream*, *The Two Gentlemen of Verona*, *As You Like It*)의 감정 패턴과 비교하여 어떤 뚜렷한 차이점이 있는지 살펴본다. 셋째, 본 연구에 사용된 감정 분석 사전인 VADER와 기존 연구의 감정 분석 사전인 AFINN 중 어떤 감정 사전이 감정 분석에 더 적합한가? 사랑 비극의 극의 흐름에 따른 남녀 주인공의 감정 변화의 선그래프 비교를 통해 살펴보고 이를 특정 대사의 누적 값과 해당하는 단어들의 점수를 비교하여 살펴본다. 넷째, 극의 흐름에 따른 남녀 주인공의 감정 변화와 주요 등장인물의 긍정/부정 감정의 변화도 설명 할 수 있는가? 남녀 주인공의 대화코퍼스를 통한 막 별 감정 분석과 주요 등장인물 간의 막 별 감정 변화의 시간시각화와 등장인물 간 연결 관계 네트워크의 관계시각화를 통해 살펴본다.

# 제 1 장 서 론

## 1.1 연구 목적 및 배경

최근 우리나라의 산업기술이 세계 최고 수준을 얻거나 유지하기 위해서는 IT 기술과 인문학의 융복합이 필요하다는 것이 당연시되고 새로운 가치 창출에 대한 기대와 관심이 증가하고 있다. 본 연구는 코퍼스 언어학 연구방법론과 네트워크 분석의 인문사회학과 자연언어처리 IT의 이공학의 융복합 연구로써, 계량 통계학의 코퍼스언어학 연구방법과 사회연결 망 프로그램을 활용한 언어네트워크 분석을 셰익스피어 비극에 접목시켜 등장인물 간의 대화코퍼스가 갖는 구조적인 특징을 시각적으로 보여줄 것이다. 특히, 셰익스피어 비평가들과 연구자들로부터 이미 검증된 주제에 대한 코퍼스 방법과 네트워크 분석 및 감정분석의 새로운 적용이라는 연구 방법적 창의성에 연구의 의미가 있다고 생각한다. 더 나아가서 본 연구의 과정들이 일반화 과정을 거치면서 셰익스피어 작품을 쉽게 이해할 수 있을 것이며, 셰익스피어의 작품뿐만 아니라 다른 작품들도 쉽게 이해할 수 있게 될 것이다.

본 연구는 셰익스피어의 비극 작품들 중 세 편의 사랑 비극 (*Romeo and Juliet*, *Othello*, *the Moor of Venice*, *Anthony and Cleopatra*)에서 주요 등장인물의 쌍방향 대화코퍼스를 구축하여 작품 및 등장인물 간의 감정분석과 네트워크 분석을 통하여 비극 작품 및 등장인물 간의 특징을 기술하고 그 특징을 시각화 하는데 연구목표를 둔다. 또한 극의 흐름에서 감정 변화를 살펴보기 위해 등장인물 간의 대화를 순서대로 번호를 매겨서 감정 분석을 실시하여 선그래프와 막대그래프로 그렸다. 첫째로 시간의 흐름에 따라 사랑 비극의 모든 등장인물의 누적된 감정값의 선그래프를 통해 사랑 비극의 특징을 나타내는 감정 패턴

을 찾고자한다. 이러한 감정 패턴이 사랑 비극에서만 나타나는 특징인지를 확인하기 위하여 셰익스피어의 희극 작품 중 세 편의 사랑 희극 (*A Midsummer Night's Dream*, *The Two Gentlemen of Verona*, *As You Like It*)의 모든 등장인물의 누적 감정값으로 나타낸 선그래프와 비교한다. 또한 비극 작품의 주요 등장인물 대화코퍼스를 기반으로 데이터를 막(Act)별로 감정 분석을 하여 주요 등장인물들의 긍정 및 부정의 감정 변화를 막대그래프로 그려 살펴보았다.

원시자료의 수집 및 가공을 위해, 본 연구의 범위인 세 편의 비극작품과 세 편의 희극작품을 현대어 버전으로 대화코퍼스를 구축하였다. 현대어 버전은 Spark Notes (<http://nfs.sparknotes.com>)에서 가져와 사용하였다. 이번 연구에서 사용된 세 편의 비극작품을 사랑 비극이라고 분류하고 목록은 *Romeo and Juliet* (1594), *Othello, the Moor of Venice* (1604), *Anthony and Cleopatra* (1606)이고, 세 편의 희극 작품은 사랑 희극이라 분류하고 목록은 *The Two Gentlemen of Verona* (1589), *A Midsummer Night's Dream* (1595), *As You Like It* (1599)이다. 모든 데이터는 데이터 전처리 과정을 거쳤으며 서로 다른 크기의 비극, 희극의 비교를 위해서 정규화하여 그래프로 그린다. 선택된 셰익스피어 작품에서 전처리 과정에서 대화코퍼스에 부합하는 대사를 선정하기 위해 독백 (Soliloquy), 방백 (Aside), 노래 (Song)는 배제한다. 등장인물 간의 네트워크 구조를 알아보기 위하여 방향성이 있는 대화코퍼스를 바탕으로 NetMiner 4.0 이라는 프로그램을 활용하여 네트워크 시각화한다. 예를 들어 Othello가 Desdemona에게 발화한 대화를 하나의 코퍼스로 만들고, Desdemona가 Othello에게 말한 대화를 다른 하나의 코퍼스로 만들어서 상호간의 감정 변화를 분석하는 것이다. 대화코퍼스는 시간의 흐름에 따른 감정 변화의 파악이 가능하기 때문에 등장인물 간의 감정 변화를 대사의 순서에 따라 분석한다. 본 연구에서 사용한 감정 사전이 기존의 감정 사전과 비교해서 어떤 사전이 더 적합한지도 살펴보기 위하여 감정 분석과 같은 양적인 통계 방법으로 사랑 비극의 남녀 주인공 간의 감정 추이를 살펴본다. 또한 네트워크의 시각화를 통해 주요 등장인물 간의 관계와 감정 분석으로는 발견하지 못한 주요 등장인물간의 관계가 긍정적인지 아니면 부정적인지, 얼마나 강한 정도로 연결이 되어있는지

를 살펴보고자 한다. 끝으로 본 연구의 결과는 빅데이터 분석, 4차 산업혁명으로 이어지는 디지털 기술의 발전에 영문학의 다양한 텍스트를 활용한 새로운 연구에 활용될 수 있을 것으로 기대한다. 그리고 다른 언어로 쓰인 극이나 소설과 같은 다양한 장르에서의 활용을 기대해본다.

제 1장에서는 연구 목적과 배경에 대해서 살펴보고 이번 연구의 개요를 간략하게 살펴본다. 제 2장에서는 이번 연구에서 중점적으로 다룰 3가지의 중요한 개념이 등장한다. 서론에서 언급했듯이, 본 연구는 셰익스피어의 비극 작품 중에서 주요 등장인물의 쌍방향 대화코퍼스를 구축하여 등장인물 간의 감정 분석과 네트워크 분석을 통하여 비극작품 및 등장인물 간의 특징을 기술하고 그 특징을 시각화하고 연구에서 사용된 감정 사전의 우위를 비교해보는 데 연구목표를 둔다. 시각화된 네트워크를 통하여 감정분석만으로는 발견해내지 못한 주요 등장인물 간의 연결 정도와 그 연결이 긍정적인지 아니면 부정적인지를 살펴보고자 한다. 감정 분석과 네트워크 분석의 시각화를 통하여 등장인물과 작품 간 감정 변화 양상을 살펴보고 우리가 잘 알고 있는 작품의 주제와 등장인물 간의 역할과 특징을 살펴본다. 따라서 등장인물 간 대화 코퍼스를 이용한 감정분석과 주요 등장인물 간 네트워크 분석을 위한 선행연구에 대해 살펴볼 것이다. 따라서 제 2장에서는 셰익스피어 관련 선행 연구, 감정 분석과 네트워크 분석에 관한 선행연구에 대해서 살펴본다. 감정 분석의 방법으로 쓰인 AFINN 과 VADER에 대해서도 살펴보았다. AFINN은 2009년부터 2011년 동안 Finn Arup Nielsen이 직접 라벨링한 -5와 +5 사이의 정수로 평가된 영어 단어 사전으로 AFINN-111이 가장 최신 버전이며 총 2,477개의 단어를 보유하고 있다. 다음으로 VADER는 Valence Aware Dictionary for Sentiment Reasoning의 약자로 Hutto and Gilbert (2014)가 감정 분석을 위해 고효율 사전을 제작하였고 문법적, 통사적 관습을 구체화 하는 규칙을 겸비한 감정 분석 알고리즘이다. 또한 등장인물 간의 네트워크를 분석하여 시각화한다. 제 3장에서는 데이터와 방법론에 대해서 알아보려고 한다. 전체 셰익스피어 작품이 아닌 세편의 비극 작품과 세편의 희극 작품을 선택한 이유에 대해서 밝히고 전체 셰익스피어 작품들 중 세편의 비극 작품의 대화코퍼스와 세 편의 희극 작품의 대화코퍼스를

수작업으로 구축한다. 제 4장에서는 감정 분석 데이터를 그래프로 시각화하여 그 결과에 대해서 논하고자 한다. 4.1에서는 세 편의 비극 작품 각각의 감정 분석을 살펴본다. 4.2에서는 세 편의 희극 작품 각각의 감정 분석을 그래프로 살펴본다. 4.3에서는 세 편의 비극 작품과 세 편의 희극 작품의 감정 분석 그래프를 하나의 그래프로 표준화하여 살펴보고 희극과 비극의 차이를 두드러지게 나타내는 패턴이 있는지 살펴본다. 4.4에서는 세편의 비극 작품에 등장하는 대표 남성 등장인물과 여성 등장인물의 감정 변화를 AFINN 감정 사전과 VADER 감정 사전을 활용한 각각의 그래프에서 차이를 나타내는 부분의 대사의 단어 별 점수를 통해 어떤 사전이 더 정확한지 확인하고 3편의 비극 작품에 등장하는 주요 등장인물 상호간의 대화코퍼스를 통해 감정 분석을 조사한다. 제 5장에서는 3편의 비극 작품의 대화코퍼스를 기반으로 한 등장인물 간의 네트워크를 시각화하여 분석한다. 마지막으로 제 6장에서는 결과를 정리하고 추후 연구 과제에 대해서 논하도록 한다. 이어서 감사의 글과 참고문헌이 수록되어있다.

## 1.2 연구 질문

본 연구는 다음의 네 가지 연구 질문의 답을 찾으려고 한다. 첫째, 사랑을 주제로 한 비극 대화코퍼스의 감정 분석을 통해 비극의 특징을 보여주는 감정 패턴을 발견 할 수 있는가? 선그래프의 시각화를 통하여 관찰할 수 있는 패턴이 몇 가지가 있는지를 고려한다. 둘째, 감정 분석을 통해 비극과 희극의 차이점을 드러낼 수 있는 패턴 또는 단서가 있는가? 관찰된 사랑 비극의 감정 패턴이 사랑 희극의 감정 패턴과는 어떤 뚜렷한 차이점이 있는지 살펴본다. 셋째, 본 연구에서 사용된 감정 분석 사전인 VADER와 기존 연구의 감정 분석 사전인 AFINN 중 어떤 감정 사전이 감정 분석에 더 적합한가? 사랑 비극의 극의 흐름에 따른 남녀 주인공의 감정 변화의 선그래프 비교를 통해 살펴보고 이를 특정 대사의 누적 값과 해당하는 단어들의 점수를 비교하여 살펴본다. 넷째, 극의 흐름에 따른 남녀 주인공의 감정 변화와 주요 등장인물의 긍정/부정 감정의 변화도 설명 할 수 있는가? 남녀 주인공의 대화코퍼스를 통한 막 별 감정 분석과

주요 등장인물 간의 막 별 감정 변화의 시간시각화와 등장인물 간 연결 관계 네트워크의 관계시각화를 통해 살펴본다.



## 제 2 장 선행 연구

지금까지 여러 가지 방법론으로 셰익스피어 관련 코퍼스언어학에 대한 다양한 연구가 진행되어왔다. 그리고 감정 분석과 네트워크 분석에 관한 선행연구들도 진행되어왔다. 제 2장에서는 본 연구에 앞서 참고하여 읽으면 도움이 될 만한 선행 연구들을 조사하였다.

### 2.1 코퍼스언어학과 셰익스피어 극

본 연구에서 사용된 사랑 비극 세 편과 사랑 희극 세 편은 Archer, Culpeper and Rayson (2005)의 분류에 따랐다. Archer et al. (2005)은 ‘사랑’이라는 주제를 가진 셰익스피어의 작품 중 희극과 비극을 통해 다른 장르에서 어떻게 ‘사랑’이라는 주제를 비유적으로 표현하는지에 대해 연구하였다. 키워드 분석 (Keyword analysis)을 통해 UCREL의 의미주석구성표 (Semantic Annotation Scheme)를 사용하여 세 편의 비극 작품 (*Romeo and Juliet*, *Othello*, *the Moor of Venice*, *Anthony and Cleopatra*)과 세 편의 희극 작품 (*A Midsummer Night's Dream*, *The Two Gentlemen of Verona* and *As You Like It*)에서 ‘사랑’이라는 개념을 탐구하였다.

Archer et al. (2005)에서 사랑 비극과 사랑 희극에 등장하는 ‘사랑’이라는 주제를 비유적으로 표현한 단어들을 나열하고 그 빈도수에 대해서 나타낸 표는 아래 <Fig. 1>과 같다.

## Comedies

Adoration (1), adore (2), adored (1), affection (7), affections (1), after-love (1), amorous (1), applaud (2), applause (1), apple of his eye (1), beloved (8), chastity (3), cherish (1), cherished (2), copulation (1), couples (4), dear (32), deflowered (1), dote (8), dotes (4), doting (1), enamoured (2), enjoy (2), fall in love (4), fallen in love (1), falling in love (1), fancies (1), fancy (7), fell in love (1), fond (7), gone for (1), impress (1), in love (34), kiss (20), kissed (3), kisses (2), kissing (3), like (117), liked (2), liking (2), love (354), loved (28), lover (32), lovers (26), loves (26), loving (10), paramour (2), precious (5), prized (1), relish (2), revelling (1), revels (5), savours (3), seduced (1), sensual (1), suitor (2), take to (1), that way (1), virgin (4), wanton (5)

## Tragedies

adore (1), adores (1), affection (8), affections (7), affinity (1), amorous (6), applauding (1), applause (1), beloved (5), believing (1), bewhored (1), camel (1), chamberers (1), chastity (3), cherish (1), cherished (1), cherishing (1), courts (1), cuckold (6), darling (1), darlings (1), dear (59), deflowered (1), devotion (4), dote (1), dotes (2), doting (5), enamoured (1), enjoy (3), enjoyed (1), fall in love (1), fancies (2), fancy (4), fond (8), impress (1), in love (8), kiss (25), kissed (6), kisses (13), kissing (6), liked (1), likes (1), liking (1), likings (1), love (259), loved (21), lover (6), lovers (9), loves (29), loving (12), lust (10), lusts (1), paramour (1), precious (6), prized (1), rate (3), rated (1), relish (1), revel (3), revels (4), sluttish (1), suitor (2), suitors (2), take to (1), that way (1), the other way (2), wantons (1), wooer (1)

<Fig. 1> Love-related lexical items occurring in the comedies and tragedies

‘사랑’이라는 주제는 Shakespeare의 작품 주제 중에서 흔하게 사용되는 주제이다. Love Tragedies는 *war*, *lack of life things*, *religion*, *the supernatural*, *lack of power*, *movement*와 같은 주요 단어가 사용되었고 Love Comedies에는 *intimate relationships*, *liking*과 같은 주요 단어가 사용되었다. 따라서 Love Tragedies서는 이상적이고 로맨틱한 의미의 ‘사랑’ 보다는 좀 더 어두운 의미로 ‘사랑’이라는 주제가 사용되었다.

그 외에도 국내외 많은 코퍼스 언어학자들은 셰익스피어 극의 텍스트 분석을 위하여 코퍼스언어학 방법론을 사용하였다. 국외의 선행 연구를 살펴보자. Busse (2002)는 저자가 쓴 문체의 기초가 되는 요소에 대해 설명하기 위해 셰익스피어 코퍼스에서 2인칭 대명사에 대하여 통사형태론적 다양성을 조사하여 2인칭 대명사의 고어형태인 ‘thou’, ‘thy’, ‘thine’, ‘ye’ 와의 차이를 분석하였다. Hota, Argamon and Chung (2006)은 셰익스피어의 극에서 등장인물의 성별에 따른 작가의 단어 선택을 통해 성별 구분에 대해 연구하였다. Scott and Tribble (2006)은 셰익스피어의 비극 작품인 *Romeo and Juliet*의 전체 텍스트에 대해 키워드 분석을 하여 설명하였다. Mahlberg (2007)는 전통적인 분석 방법을 사용하는 문학 비평가들이 이전의 방법으로는 알아차리기 어려웠던 문학적 관심을 높이기 위해 코퍼스 분석 방법에 대해 더 많은 연구를 해야 한다고 설명했다. Culpeper (2009)는 텍스트 내에서 키워드, 품사, 주요 의미론적 범주 그리고 그것들의 역할에 대한 탐구하였다. 셰익스피어의 비극 작품인 *Romeo and*

*Juliet*을 통해 첫 번째로, 키워드는 스타일 마커와 동일한 역할을 하며 이를 통해 세 가지 종류 (사람과 사람에 관한 키워드, 텍스트 내 키워드, 관념적인 키워드)의 키워드에 대해 식별 할 수 있다고 설명하였고 두 번째로, 키워드뿐만 아니라 주요 품사, 주요 의미론적 도메인의 분석을 통해 선택한 텍스트의 특징을 직관에 의존 하지 않고 어휘적, 문법적 패턴을 통해 특징을 발견해 낼 수 있다고 발표했다.

다음으로 국내의 선행연구를 살펴보자. 하명정 (2013)은 코퍼스언어학의 분석 도구인 WordSmith를 이용하여 셰익스피어의 비극작품인 *Romeo and Juliet*의 전체 텍스트에 대한 문체적 특성과 주요테마를 조사하기 위하여 키워드 분석을 하였다. 장세은, 이수상, 송원문, 정해룡, 이성민, 김재훈 (2017)은 셰익스피어를 주제로 하는 국내외 학술논문 영어 초록을 모아서 구성한 코퍼스 Shakespeare Research Article English Abstract Corpus (이하, SRAEAC)를 기계학습 알고리즘 기반의 토픽모델링을 적용하여 SRAEAC에서 자동으로 추출된 토픽을 Scott (2009)의 키워드 분석과 비교하여 통시적으로 셰익스피어 연구에서 어떤 기간에 어떤 토픽이 있었는지 또한 시간의 흐름에 따른 토픽의 변화가 어떠한지를 알아보하고자하였다. Liu and Jhang (2018)은 셰익스피어의 작품들 중 ‘사랑’을 주제로 한 비극 작품 세 편의 그룹과 희극 작품 세 편의 그룹에 대해 키워드와 키클러스터를 조사하였다. 사랑을 주제로 한 비극과 희극의 상위 20개의 키워드를 살펴본 결과, 여섯 편의 작품은 1위에서 6위까지의 사랑 관련 단어들을 공유하고 있으며, 두 그룹의 ‘사랑’이라는 단어에 대한 분산 정도를 조사하였을 때 사랑 희극 그룹이 더 높은 분산도를 가졌다고 설명했다. 또한 사랑 희극 그룹에서는 ‘사랑’과 관련된 *love, lover, lovers, sweet, wit, love’s*와 같이 긍정적인 단어들 많이 발견된 반면 사랑 비극 그룹에서는 *dead, poison, beguil’d, death, soldier’s*와 같이 부정적인 단어들 많이 발견되었다. 이러한 키워드와 키클러스터 분석을 통해 비극과 희극을 구별 할 수 있다고 기술했다. 이경은, 장세은 (2018)은 셰익스피어 비극 작품의 주요 등장인물의 대화코퍼스 분석을 통한 페미니즘에 대한 연구가 있었다. 키워드 분석(keyword analysis)을 통해 셰익스피어의 10대 비극 작품 (*Titus Andronicus,*

*Romeo and Juliet, Julius Caesar, Hamlet, Othello, King Lear, Macbeth, Anthony and Cleopatra, Timon of Athens, Coriolanus*)에 등장하는 남성 등장인물과 여성 등장인물들의 대화코퍼스의 남성성과 여성성을 나타내는 단어를 살펴보고 연관 있는 단어 또는 함께 등장하는 단어들에 대해 논의하였다.

## 2.2 감정 분석

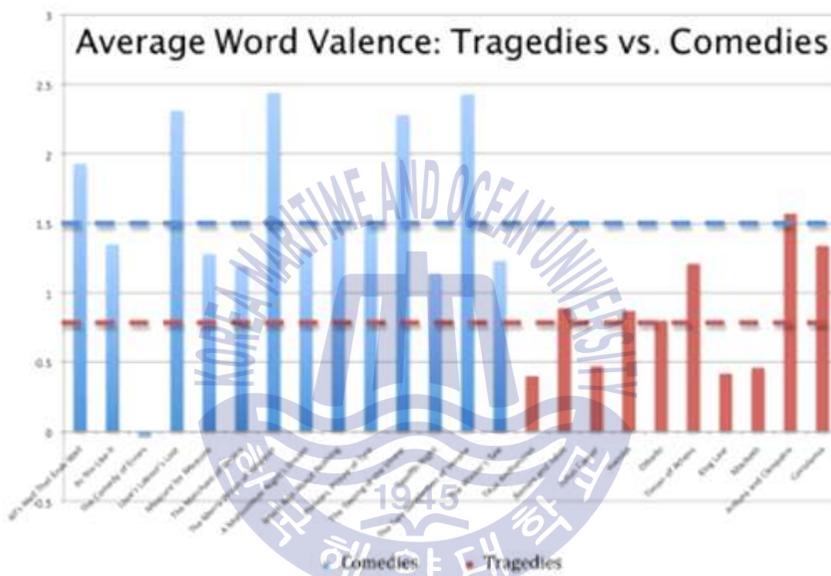
감정 분석 (sentiment analysis)이란 텍스트에 나타난 사람들의 태도, 의견, 성향과 같은 주관적인 데이터를 분석하는 자연어처리기술이다. 이는 ‘오피니언 마이닝 (opinion mining)’으로도 불린다. 다시 말해 텍스트에 나타난 성향이 긍정적인지, 부정적인지 혹은 중립적인지를 결정하기 위해 쓰이는 방법 중 하나이다.

셰익스피어 극의 대화코퍼스를 활용한 국외의 선행연구는 최근 2000년대부터 시작하였으나 국내 연구는 최근 2018년 초에 시작되었다. 우선 국외 선행연구를 살펴보자. Stiller, Nettle and Dunbar (2003)는 셰익스피어 작품 내 인물 간 연결에 따른 네트워크에 관한 연구이다. 작품 내 등장인물 수가 증가할수록 등장인물 간 연결되는 정도가 현저히 줄어든다는 연구 결과를 발표했다. Mohammad (2011)는 또한 셰익스피어의 비극에서 *Hamlet*과 희극에서 *As You Like It*을 감정 분석을 효과적인 시각화와 함께 사용하여 개별적인 데이터의 감정을 정량화하고 추적하는 방법에 대해 설명하였다. 그리고 비극인 *Hamlet*, 희극인 *As You Like It*, 공포 소설인 *Frankenstein*을 비교하여 소설의 장르에 따른 감정의 단어들을 살펴보았다.

본 연구에서 사용된 VADER 감정 사전과 비교하여 기존 연구에서 사용된 AFINN 감정 사전에 관한 선행연구에 대해서 살펴보자. Nalisnick and Baird (2013)은 셰익스피어의 작품에서 등장인물간의 감정 분석을 한 연구가 진행되었다. 본 연구의 방법과 마찬가지로 등장인물들의 대화코퍼스를 대중적인 감정 사전인 AFINN을 활용하여 도출해낸 값들의 합으로 등장인물간의 감정 분석을

시행하였다. 감정 분석은 단어들이 감정 점수 (valence)가 매겨지는데 이때 감정 점수는 긍정적 또는 부정적으로 평가된 일련의 감정들을 표현하기 위한 점수이다.

비극 작품과 희극 작품을 AFINN을 활용하여 단어의 누적 감정 점수를 평균 낸 그래프는 아래 <Fig. 2>와 같다.



<Fig. 2> Average Word Valence : Tragedies vs Comedies

위 <Fig. 2>는 셰익스피어의 14개의 희극 작품과 10개의 비극 작품을 AFINN 감정 사전을 활용하여 도출해낸 감정 점수의 누적 값을 평균 내어 막대그래프로 나타내었다. 다시 말해 비극 작품과 희극 작품의 감정 단어의 누적 감정 점수의 평균을 비교하였다.

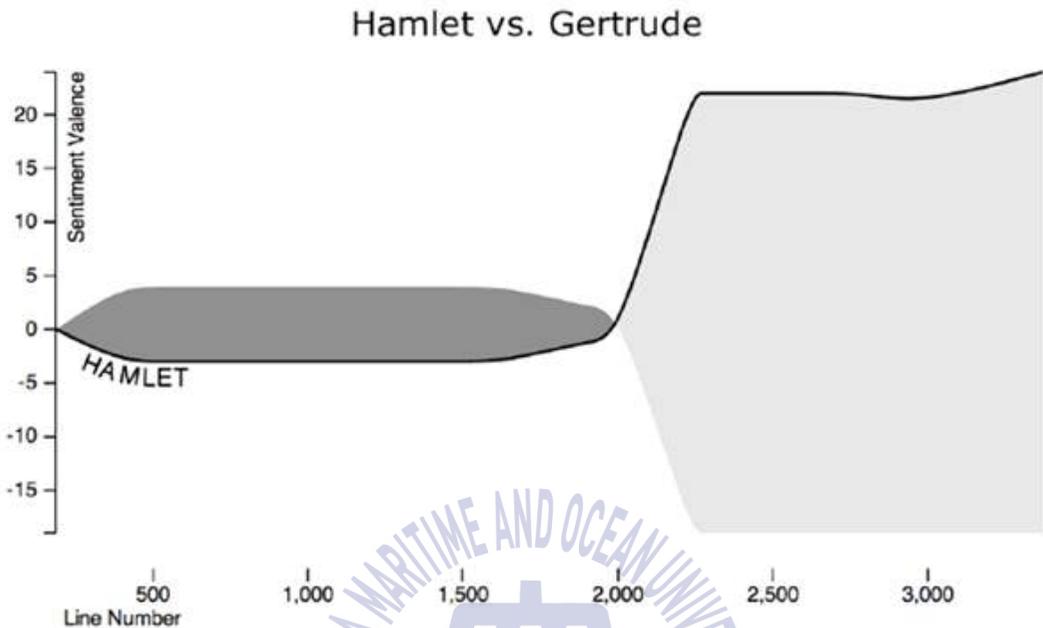
기존의 연구에서 Hamlet을 제외한 나머지 인물들에게서 도출해낸 감정값의 표는 아래 <Fig. 3>으로 나타냈다.

Character	Hamlet's Sentiment Valence Sum
Guildestern	31
Polonius	25
Gertrude	24
Horatio	12
Ghost	8
Marcellus	7
Osric	7
Bernardo	2
Laertes	-1
Ophelia	-5
Rosencrantz	-12
Claudius	-27

<Fig. 3> Hamlet's Sentiment Valence Sum

위 <Fig. 3>에서는 Hamlet 자체의 감정값을 알 수 없어서 나머지 인물들의 감정값을 Hamlet과 비교해 볼 수 없다는 점이 아쉽다. 가장 부정적인 감정을 가진 인물은 Claudius로 -27이라는 값을 가진다. 그리고 극의 전개상 Hamlet은 자신의 아버지인 King Hamlet의 죽음에 연관이 되어있다고 생각한 어머니 Gertrude와 부정적인 감정을 가질 것이라고 예상했다. 그러나 의외로 Gertrude의 AFINN 값은 +24이다.

작품 *Hamlet, Prince of Denmark*에서 Hamlet과 Gertrude의 AFINN값을 비교하여 <Fig. 4>에서 각 주인공의 감정 변화를 그래프로 나타냈다.



<Fig. 4> Comparing between Sentiment Valences of Hamlet-Gertrude

위 <Fig. 4>에서 Gertrude에 대한 Hamlet의 감정 그래프는 검은색 실선이고, Hamlet에 대한 Gertrude의 감정 그래프는 진회색으로 표시된 부분이다. 두 등장인물의 감정이 교차하는 부분이 3막 5절: The Queen's Closet에 해당한다고 설명했다.

다음으로 국내 선행연구를 살펴보자. 이경은, 장세은, 김재훈, 박호민(2018)은 셰익스피어의 작품 중 사랑을 주제로 하는 비극 작품 세 편과 희극 작품 세 편의 등장인물들의 대화코퍼스를 VADER감정 분석 알고리즘을 이용하여 감정분석하고 그 결과를 시각화 하였다. 먼저 사랑 비극 세 편과 사랑 희극 세 편의 전체 대화코퍼스를 통해 감정 분석하여 비극과 희극을 구분 짓는 패턴을 찾아 보고 주요 남성 등장인물과 여성 등장인물의 대화코퍼스를 VADER 감정 분석 사전과 AFINN 감정 분석 사전의 결과를 비교하였다. 막 별 주요 등장인물들의 감정 변화와 긍정/부정 감정 관계에 대하여 살펴보고 결과를 NetMiner 4.3을

사용하여 시각화하였다. 장세은, 김재훈, 박호민 (2018)은 VADER 감정 분석 알고리즘과 셰익스피어의 10편의 비극 작품의 대화코퍼스를 사용하여 막을 기준으로 작품 및 주요 등장인물의 감정의 변화와 추이변화 양상을 관찰하여 분석하였다. 막 별 10편의 비극 작품의 감정 추이변화를 관찰하여 마지막 5막 마지막 부분의 누적 감정 그래프의 모양이 모두 하강 하지 않는다는 것을 발견하였고 셰익스피어의 4대 비극은 모두 2막 중간쯤에서 누적 감정값이 교차하는 그래프 모양을 보인다는 사실을 발견하였다. 또한 주요 등장인물 간의 감정의 변화의 비교에서 남녀주인공의 감정의 변화를 시각적으로 보여주었다. 막 별로 주요 등장인물 간의 비교를 통해 등장인물 간의 감정을 비교하였고 특정 등장인물의 극의 흐름에 따른 감정 변화를 살펴보았다.)

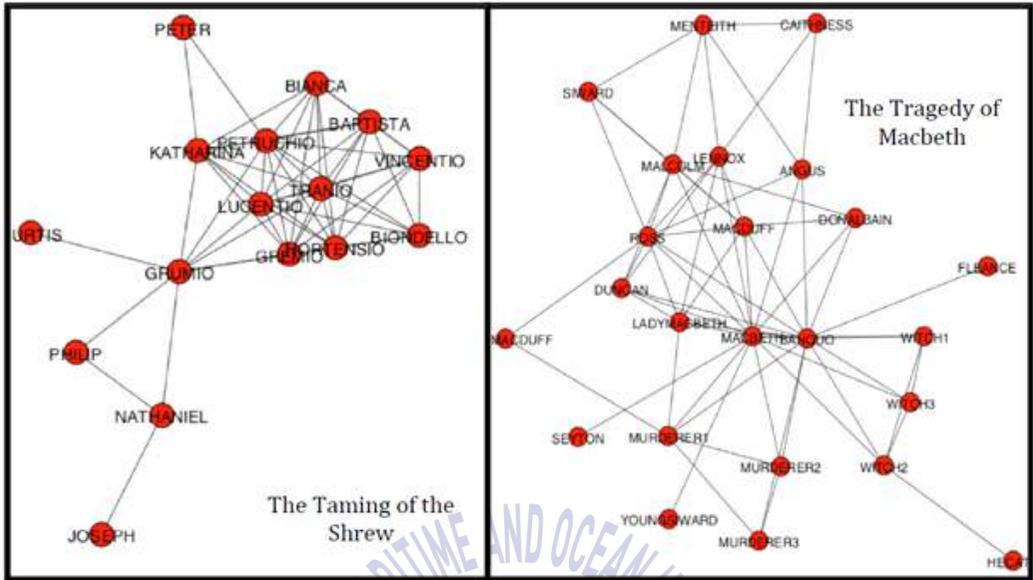
## 2.3 네트워크 분석



1) 셰익스피어 극의 작품 분석뿐만 아니라 다른 분야에서도 감정 분석을 사용한 선행 연구를 살펴보자. Laurel Hart (2013)는 일반적으로 감정 분석으로 주어진 텍스트의 감정을 자동으로 탐지하고 그것을 특성화 할 수 있다고 발표했다. 또한 텍스트 내에서 풍자 및 희화화된 영역을 탐지하고, 감정을 분류해 내고, 그 감정의 극성(긍정/부정)을 분류해 내는 것은 아주 작은 부분에 지나지 않으며, 그 자체가 전산 언어학 분야의 아주 작은 부분이라고 언급했다. 더 나아가 사람이 감지해 내지 못하는 부분까지도 감지 할 수 있을 것이며 이를 AI 기술의 발전을 도모할 수 있을 것이라는 연구 결과를 발표했다. Modha, Pandi and Modha (2013)는 웹에 존재하는 비정형 데이터에 대한 감정 분석을 수행하기 위한 기존의 방법들 및 접근법에 대해 논의가 이루어졌는데, 현재까지 연구에 활용되고 있는 감정 분석은 객관적인 기술 보다 주관적인 기술이나 주관에 초점이 맞춰져 있다고 발표했다. 따라서 그들은 감정 분석을 위한 객관적인 기술뿐만 아니라 주관적인 기술을 분류하고 다루는 새로운 접근법을 제안했다. 그들은 분류를 위해 네 가지 단계를 소개했다. Step 1) 우선, 주관적이든 객관적이든 간에, 문서의 문장을 개방형 또는 비개방형, 두 개의 범주로 분류한다. Step 2) 만일 개방형 문장이라면, 그들이 주관적인지 객관적인지를 분류해 낸다. Step 3) 주관적인 문장을 긍정적, 부정적 또는 중립적인 범주로 분류하는 단계이다. 복잡한 유형의 문장을 위해서는 문맥이나 의미지향성을 살펴볼 필요가 있다. Step 4) 네 번째 단계는 객관적인 문장을 긍정적, 부정적 또는 중립적인 범주로 분류한다. 마찬가지로 필요할 때에는 문맥이나 의미지향성을 살펴볼 필요가 있다고 설명했다. 또한 감정 분석은 언어학뿐만 아니라 e-commerce분야에서도 활발하게 활용되고 있다. 일반적으로 NLP 기술은 소비자들이 작성한 리뷰에서 제품의 특징을 도출해 내는 용도로 사용되어 왔다. Hussain, Sattar, Afzal (2015)은 감정 분석을 기반으로 하여 데이터의 특징을 도출해 내는 방법을 설명했다. 감정 분석을 통해 데이터의 특징을 찾아낼 때 가장 중요한 과제는 찾아낸 특징들을 세분화하는 것이라고 연구를 통해 밝혔다.

기존의 셰익스피어 작품의 네트워크 분석 연구는 셰익스피어 작품에 등장하는 주요 등장인물 또는 모든 등장인물 간의 연결을 통하여 네트워크의 구조와 연결강도를 시각화하여 작품 간의 특징을 살펴보려 하였다. 문학작품의 네트워크 분석은 최근 21세기 초부터 시작되었다. Stiller and Hudson (2005)은 셰익스피어의 10편의 작품의 네트워크 분석을 통해 작품 속 구조적 특징을 살펴보고 각 등장인물 간의 연결된 수와 사람이 관찰하였을 때 발견할 수 있는 네트워크 연결의 수를 비교하였다. 10편의 셰익스피어 작품의 평균 연결 정도를 표로 정리하여 보여주었다. Moretti (2011)는 네트워크 분석은 극의 내용을 보기 쉽게 시각화하여 컴퓨터를 사용한 문학의 분석 방법에서 미처 발견하지 못한 부분을 쉽게 발견하기에 적합하다고 주장하였다.

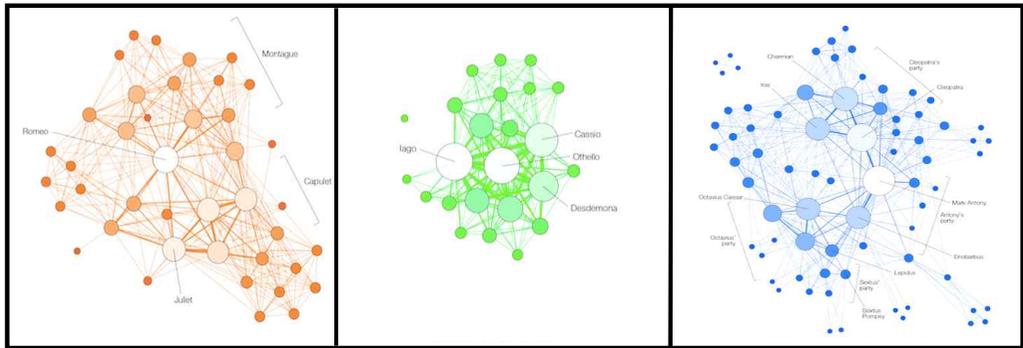
본 연구의 감정 네트워크 분석의 방법론은 Nalisnick (2013)에서 착안하였다. Nalisnick (2013)은 사전 기반 알고리즘인 AFINN을 활용한 셰익스피어 작품의 감정 분석 연구와 더불어 작품 내 인물 간 네트워크를 생성하여 PageRank 알고리즘을 응용하여 주요 캐릭터를 검토하고 기계학습 알고리즘을 사용하여 네트워크 분석을 시도하였다. 나아가 Nalisnick and Baird (2013a, 2013b)는 사전 기반 알고리즘은 AFINN 감정 사전을 활용하여 셰익스피어 작품의 감정 분석을 연구하였을 뿐만 아니라 감정 분석을 소셜 네트워크와 결합시켜 연구하였다. *The Taming of the Shrew*와 *The Tragedy of Macbeth*를 소셜 네트워크로 연결시켜 나타낸 아래 <Fig. 5>로 나타내었다.



<Fig. 5> The Social Networks for *The Taming of the Shrew* and *The Tragedy of Macbeth*

위 <Fig. 5>에서 보는 바와 같이 *The Taming of the Shrew*는 *The Tragedy of Macbeth*보다 등장인물의 수는 적지만 등장 인물간의 연결 정도는 더 많은 것을 확인 할 수 있다.

최근에 Grandjean (2015)는 셰익스피어의 11개의 비극 작품의 모든 등장인물간의 네트워크 구조의 밀집도를 %로 나타내었다. 네트워크 밀집도는 등장인물간 얼마나 밀집해 있는지를 측정한 것이다. 본 연구의 범위인 세편의 사랑 비극의 네트워크만 첨부한 것은 아래 <Fig. 6>과 같다.



<Fig. 6> Network Structures of Three Love Tragedies

위 <Fig. 6>에서 보는 바와 같이 세 편의 비극의 *Romeo and Juliet*, *Othello*, *the Moor of Venice*, *Antony and Cleopatra*를 연도 순서대로 나열하였다. 사랑 비극 세 편의 네트워크의 구조에서 밀집도는 *Romeo and Juliet*은 37%, *Othello, the Moor of Venice*에서는 55%, *Antony and Cleopatra*에서는 17%의 밀집도를 보였다.

## 2.4 감정 사전

여기서는 기존에 활용되었던 AFINN 감정 사전과 본 연구에서 활용된 VADER 감정 사전에 대해서 기술한다.

### 2.4.1 AFINN

AFINN은 2009년에서 2011년 동안 Nielsen (2011)이 직접 값을 매겨 가장 부정적인 단어는 -5, 가장 긍정적인 단어는 +5의 값으로 평가된 영어 단어 감정 사전이다. 아래 <Fig. 7>은 AFINN 감정 사전의 일부를 발췌한 것이다.

accident	-2
accidental	-2
accidentally	-2
accidents	-2
accomplish	2
accomplished	2
accomplishes	2
accusation	-2
accusations	-2

<Fig. 7> AFINN Sentiment Dictionary

위의 <Fig. 7>에서 보이는 바와 같이 단어 마다 점수가 하나의 점수가 매겨져 있는 것을 확인 할 수 있다. AFINN-111이 가장 최신 버전이며 총 2,477개의 단어를 보유하고 있다. 당시 연구적으로 가장 인정받던 ANEW라는 감정 사전의 한계를 극복하기 위해 제작되었다. 기존 감정 사전들이 마이크로 블로그에서 주목하지 않았던 이모티콘, 약어, 슬랭에 주목했다는 점에서 높은 의의를 가진다.

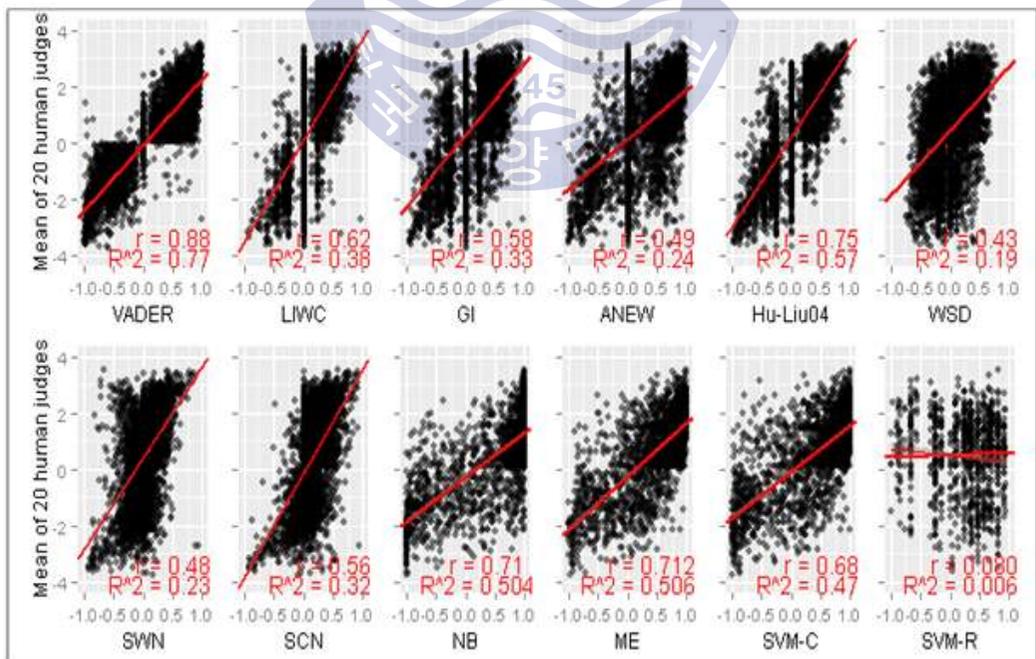
#### 2.4.2 VADER

Hutto and Gilbert (2014)가 기존의 AFINN보다 감정 분석을 위한 고효율의 사전 Valence Aware Dictionary for Sentiment Reasoning(이하, VADER) 감정의 긍정 및 부정의 양극성과 민감도 (강도)를 바탕으로 제작한 사전이다. 이는 전문가들이 미리 설정해 놓은 점수를 사용하여 양적 분석과 그에 따른 양적 검증 을 결합한 사람 중심적 접근법이기 때문에 텍스트 감정 분석에 주로 사용되는 모델이다. VADER 감정 사전의 일부를 발췌하여 아래 <Fig. 8>로 나타내었다.

accident	-2.1	0.83066	[-2, -2, -1, -3, -4, -2, -2, -1, -2, -2]
accidental	-0.3	0.45826	[-1, -1, 0, 0, 0, 0, 0, 0, -1, 0]
accidentally	-1.4	0.91652	[-2, 0, -2, 0, -3, -1, -1, -1, -2, -2]
accidents	-1.3	0.78102	[-1, -1, -1, -1, -2, 0, -3, -1, -2, -1]
accomplish	1.8	0.6	[1, 2, 3, 2, 2, 2, 1, 1, 2, 2]
accomplished	1.9	0.53852	[2, 2, 2, 1, 2, 2, 3, 1, 2, 2]
accomplishes	1.7	0.9	[2, 2, 1, 0, 2, 3, 3, 1, 1, 2]
accusation	-1.0	1.09545	[-1, -1, -2, -2, -2, -1, -1, -1, 2, -1]
accusations	-1.3	1.26886	[-2, -2, -1, -3, -2, -1, -1, 2, -2, -1]

<Fig. 8> VADER Sentiment Dictionary

위 <Fig. 8>에서 보이는 바와 같이 각 단어 마다 개별적으로 점수가 매겨지며 10명의 전문가가 매긴 점수를 평균을 낸 것이다. 기존 마이크로 블로그 감정 분석을 위한 사전들과는 달리 도메인에 제약되지 않는 감정 사전으로 기존에 배포되었던 사전에서 확장되고 검증된 인원들의 의견이 반영되었다.



<Fig. 9> Sentiment Scores Form VADER and 11 Other Sentiment Analysis Tools

VADER는 문법적, 통사적 관습을 구체화 하는 규칙을 겸비한 감정 분석 알고리즘으로 아래 <Fig. 9>에서 보는 바와 같이 기존 사전 및 기계 학습 알고리즘(총 11가지)들과 비교했을 때 가장 우수한 성과를 냈으며 특정분야에서는 사람의 정확도와 동일 수준이다(Hutto and Gilbert, 2014).

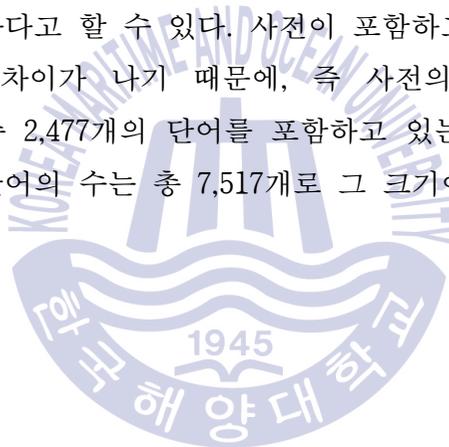
위의 <Fig. 9>는 4K가 넘는 트윗에 대해 VADER 알고리즘과 11개의 다른 감정 분석 도구와 알고리즘들의 정확도를 나타내는 그림이다. 총 11가지의 기계 학습 알고리즘에 대해 각 알고리즘에 대해 설정해놓은 값들에 어떤 방법의 알고리즘이 가장 설정 해놓은 기준치에 가까운 값을 나타내는지를 도표로 나타냈다. 그중에서 VADER 알고리즘이 가장 가까운 값을 보여주었다. 각 분포도는 4개의 부분으로 나뉘는데 왼쪽 아래는 참 부정 값을 나타내고, 오른쪽 위는 참 긍정 값을 나타낸다. 나머지 왼쪽 위는 거짓 부정 값을, 오른쪽 아래는 거짓 긍정 값을 나타낸다.

VADER 알고리즘 외 7가지 다른 알고리즘들의 정확도를 사람의 정확도와 비교하여 아래 <Fig. 10>의 표로 정리하여 나타내었다.

	Correlation to ground truth (mean of 20 human raters)	3-class (positive, negative, neutral) Classification Accuracy Metrics			Ordinal Rank (by F1)		Correlation to ground truth (mean of 20 human raters)	3-class (positive, negative, neutral) Classification Accuracy Metrics		
		Overall Precision	Overall Recall	Overall F1 score				Overall Precision	Overall Recall	Overall F1 score
<b>Social Media Text (4,200 Tweets)</b>						<b>Movie Reviews (10,605 review snippets)</b>				
Ind. Humans	0.888	0.95	0.76	0.84	2	1	0.899	0.95	0.90	0.92
VADER	0.881	<b>0.99</b>	<b>0.94</b>	<b>0.96</b>	1*	2	0.451	0.70	0.55	0.61
Hu-Liu04	0.756	0.94	0.66	0.77	3	3	0.416	0.66	0.56	0.59
SCN	0.568	0.81	0.75	0.75	4	7	0.210	0.60	0.53	0.44
GI	0.580	0.84	0.58	0.69	5	5	0.343	0.66	0.50	0.55
SWN	0.488	0.75	0.62	0.67	6	4	0.251	0.60	0.55	0.57
LIWC	0.622	0.94	0.48	0.63	7	9	0.152	0.61	0.22	0.31
ANEW	0.492	0.83	0.48	0.60	8	8	0.156	0.57	0.36	0.40
WSD	0.438	0.70	0.49	0.56	9	6	0.349	0.58	0.50	0.52
<b>Amazon.com Product Reviews (3,708 review snippets)</b>						<b>NY Times Editorials (5,190 article snippets)</b>				
Ind. Humans	0.911	0.94	0.80	0.85	1	1	0.745	0.87	0.55	0.65
VADER	0.565	0.78	0.55	0.63	2	2	0.492	0.69	0.49	0.55
Hu-Liu04	0.571	0.74	0.56	0.62	3	3	0.487	0.70	0.45	0.52
SCN	0.316	0.64	0.60	0.51	7	7	0.252	0.62	0.47	0.38
GI	0.365	0.67	0.49	0.55	5	5	0.362	0.65	0.44	0.49
SWN	0.325	0.61	0.54	0.57	4	4	0.262	0.57	0.49	0.52
LIWC	0.313	0.73	0.29	0.36	9	9	0.220	0.66	0.17	0.21
ANEW	0.257	0.69	0.33	0.39	8	8	0.202	0.59	0.32	0.35
WSD	0.324	0.60	0.51	0.55	6	6	0.218	0.55	0.45	0.47

<Fig. 10> VADER Performance Compared to Human Raters

VADER 알고리즘을 이용하여 3개의 등급 (긍정, 부정, 중립)으로 분류하고 그 외 7개의 서로 다른 기존 감정 사전과 기계 학습 알고리즘에 미리 설정된 어휘 목록과 비교하였다. 정확도가 전문가들의 정확도와 얼마나 차이가 나는지 비교하고 그 순위를 매겨 보기 쉽게 나타냈다. 그리고 전문가의 정확도와 VADER 감정 사전의 정확도가 가장 비슷하다는 결과를 살펴볼 수 있다. 실질적으로 AFINN 감정 사전과 VADER 감정 사전을 비교했을 때, 가장 큰 차이는 VADER만이 가지고 있는 언어적 규칙기반 알고리즘이며 단순한 사전 기반의 기계적인 감정 분석보다 인간 중심적인 감정 분석이 가능하다. 또한 10명의 전문가들이 최대 +4, 최저 -4의 점수를 매겨 평균을 낸 VADER와 달리, AFINN은 저자 Finn 혼자 최대 +5, 최저 -5의 점수를 매겨 제작하였기 때문에 단어에 주어진 점수에 대한 검증이 부족하다고 할 수 있다. 사전이 포함하고 있는 단어와 점수 제작에 참여한 인원이 차이가 나기 때문에, 즉 사전의 크기도 차이가 난다. AFINN 감정 사전은 총 2,477개의 단어를 포함하고 있는 반면, VADER 감정 사전이 포함하고 있는 단어의 수는 총 7,517개로 그 크기에서 대략 3배정도 차이가 난다.



## 제 3 장 데이터와 방법론

3.1에서는 세 편의 사랑 비극과 세 편의 사랑 희극 각각의 일반 통계정보와 각 작품의 대화코퍼스 정보를 살펴본다. 그리고 각 작품에 등장하는 등장인물들의 일반 통계정보도 살펴보도록 한다. 3.2에서는 이번 연구에 사용된 연구 방법론인 감정 분석과 네트워크 분석에 대하여 설명한다.

### 3.1 데이터

지금부터 본 연구의 범위인 세 편의 비극 작품 (*Romeo and Juliet* (1594), *Othello, the Moor of Venice* (1604) and *Antony and Cleopatra* (1606))의 데이터를 살펴보고 그와 비교할 세 편의 희극 작품 (*The Two Gentlemen of Verona* (1589), *A Midsummer Night's Dream* (1595) and *As You Like It* (1599))의 데이터를 살펴본다.

#### 3.1.1 사랑 비극

본 연구의 범위인 세 편의 비극 작품을 현대어 버전으로 대화코퍼스를 구축하였다. 현대어 버전은 SparkNotes (<http://nfs.sparknotes.com>)에서 가져와 사용하였다. 이번 연구에서 사용된 세 편의 비극 작품을 love tragedies라고 분류하고 목록은 *Romeo and Juliet* (1594), *Othello, the Moor of Venice* (1604), *Antony and Cleopatra* (1606) 이다. 모든 데이터는 데이터 전처리 과정을 거치

며 정규화하였다. 선택된 셰익스피어 작품에서 전처리 기준을 정하기 위하여 대화코퍼스에 부합하는 대사를 선정하고 다음 독백 (Soliloquy), 방백 (Aside), 노래 (Song)은 배제하였다. 독백이란 한 사람의 인물에 의해 행하여지는 누구로부터 방해 받거나 매개되지 않는 다소 긴 발화로 인물이 스스로에게 말하는 것으로 자신의 생각이나 감정은 드러내며 이러한 것들을 관객과 공유하는 장치이다. 따라서 청자가 없으므로 대화코퍼스에서 제외되었다. 다음으로 방백은 곁에 사람을 두고도 혼자서 지껄이는 말이다. 이때 곁에 사람이 그 말을 알아듣지 못하는 경우에 방백의 효과는 살아난다. 즉, 방백은 관객을 향한 말이다. 독백과 마찬가지로 청자가 없어 대화코퍼스에서 제외되었다. 마지막으로 노래는 대화가 아니므로 삭제되었다.

각 비극 작품 세편의 일반 통계를 알아내기 위해 Wordsmith 4.0이라는 프로그램을 사용했다. 표를 설명하기에 앞서 먼저 표에 사용되는 STTR의 개념에 대해 살펴보도록 한다.

“[T]he Standardized type/token ratio (STTR) is computed every  $n$  words as Wordlist goes through each text file. By default,  $n=1,000$ . In other words the ratio is calculated for the first 1,000 running words, then calculated afresh for the next 1,000, and so on to the end of your text or corpus. A running average is computed, which means that you goes an average type/token ration based on consecutive 1,000-word chunks of text (Scott, 2009).”

위의 STTR의 영문 설명에서 보는 바와 같이 Scott (2009)은 워드리스트가 텍스트 파일에서 기본 값은  $n=1,000$ 일 때 각각의  $n$ 개의 단어로 계산된다고 설명했다. 다시 말해 STTR을 계산하기 위해 처음 1,000개의 running words를 계산하고 다음 새로운 1,000개의 running words를 계산하고 텍스트 또는 코퍼스의 마지막까지 1,000개의 단위로 계산한다. running words의 평균을 계산한다는 것은 텍스트의 1,000개의 연속체에 대해 type/token 비율의 평균을 내는 것이라고 설명했다.

본 연구의 범위인 세 편의 사랑 비극을 연도별로 정리한 통계 자료는 아래의 <Table 1>과 같다.

<Table 1> General Information of Three Love Tragedies

N	Chronology	text files	tokens (running words)	types (distinct words)	type/token ratio (TTR)	standardised TTR
		Overall	75,761	7,125	9.40	41.96
1	1595	Romeo and Juliet	24,617	3,531	14.34	41.64
2	1605	Othello, the Moor of Venice	26,601	3,631	13.65	41.02
3	1607	Anthony and Cleopatra	24,543	3,700	15.08	43.28

위의 <Table 1>은 제작연도로 순서를 매겼다. 세 작품을 크기가 큰 순서로 정리를 하면 *Anthony and Cleopatra* > *Othello, the Moor of Venice* > *Romeo and Juliet* 순이다. Token 별 크기를 정리해 보자면 *Othello, the Moor of Venice* > *Romeo and Juliet* > *Anthony and Cleopatra* 순서 이다. 다음으로 Type을 기준으로 정리해 보자면 *Anthony and Cleopatra* > *Othello, the Moor of Venice* > *Romeo and Juliet* 이다. STTR 값을 기준으로 순서를 매겨 보자면 *Anthony and Cleopatra* > *Romeo and Juliet* > *Othello, the Moor of Venice* 이다. 종합해보면 세 작품들 중 *Anthony and Cleopatra*는 전체 사용된 단어에서 다양한 단어가 많이 사용된 반면에, *Othello, the Moor of Venice*는 전체 사용된 단어 수에 비해 중복 사용된 단어의 수가 가장 많다는 것을 확인 할 수 있다.

### 3.1.1.1 *Romeo and Juliet*

극 *Romeo and Juliet*의 대화코퍼스를 정리한 표는 <Table 2>와 같다.

<Table 2> Dialogue Corpus Information of *Romeo and Juliet*

N	Act(Scene)	Speaker(s)	Dialogue	Listener(s)
1	1(1)	SAMPSON	Gregory, I swear, we can't let them humiliate us. We won't take their garbage.	GREGORY
...				
54	1(1)	PRINCE ESCALUS	Your ebels! Enemies of the peace! Men who turn their weapons against their own neighbors—They won't listen to me?—You there! You men, you beasts, who satisfy your anger with fountains of each others' blood! I'll have you tortured if you don't put down your swords and listen to your angry prince. Everyone else, go away for now.	SAMPSON GREGORY ABRAM BALTHASAR BENVOLIO TYBALT
...				
923	5(3)	PRINCE ESCALUS	We settle a dark peace this morning. The sun is too sad to show itself. Let's go, to talk about these sad things some more. Some will be pardoned, and some will be punished. There was never a story more full of pain than the story of <i>Romeo and Juliet</i> .	ALL

먼저 극 전체를 막과 장으로 정리하고 대사 순서대로 나열하여 번호를 매겼

다. 각 대사의 발화자와 청자 그리고 대사에서 언급되는 등장인물들을 표시했다. 대사에서 직접적으로 언급되거나 짐작할 수 있는 시간과 장소도 표시하였다. 또한 총 대화 수는 923으로 앞서 언급했던 바와 같이 세 작품 중 크기가 가장 작은 것이 확인된다. *Romeo and Juliet*은 청자가 최대 여섯 명까지 등장한다.

극 *Romeo and Juliet*에 등장하는 모든 등장인물들에 대한 통계 자료를 정리한 표는 아래 <Table 3>이다.

<Table 3> General Information of Characters in Romeo and Juliet

N	text files	tokens (running words)	types (distinct words)	type/token ratio (TTR)	standardised TTR
	Overall	24,617	3,531	14.34	42.27
1	ROMEO	4,808	1,322	27.50	43.80
2	JULIET	4,396	1,126	25.61	40.60
3	FRIAR LAURENCE	2,853	932	32.67	45.40
4	NURSE	2,228	624	28.01	36.90
5	CAPULET	2,205	710	32.20	40.75
6	MERCUTIO	2,142	807	37.68	45.40
7	BENVOLIO	1,204	504	41.86	44.10
8	LADY CAPULET	902	376	41.69	
		...			
34	MUSICIANS	1	1	100.00	

극 *Romeo and Juliet*에는 34명의 인물이 등장한다. 총 34명의 등장인물 가운데 각 인물이 가지는 Token이 1,000이상인 등장인물인 Romeo, Juliet, Friar Laurence(Lawrence), Nurse, Capulet, Mercutio, Benvolio만으로 주요 등장인물 감정 분석과 네트워크 분석을 시행하였다.

### 3.1.1.2 *Othello, the Moor of Venice*

극 *Othello, the Moor of Venice*의 대화코퍼스 정보를 정리한 표는 아래 <Table 4>와 같다.



<Table 4> Dialogue Corpus Information of *Othello, the Moor of Venice*

N	Act(Scene)	Speaker	Dialogue	Listener(s)
1	1(1)	RODERIGO	Come on, don' t tell me that. I don' t like it that you knew about this, Iago. All this time I' ve thought you were such a good friend that I' ve let you spend my money as if it was yours.	IAGO
...				
72	1(2)	OTHELLO	Put away your swords. They' ll get rusty in the dew. Sir, your age and status inspire more respect than your weapons do.	RODERIGO IAGO BRABANTIO
...				
1210	5(2)	LODOVICO	(to IAGO) You blood thirsty dog, you' re crueler than sadness or hunger, crueler than the sea. Look at the sedead people on this bed. You did all this. He makes me sick. Take him away. Gratiano, take care of the house, and take the Moor' s property. You' ve inherited everything. —(to CASSIO) Governor, I leave it in yourh and stop unish this evil villain: just decide the time, the place and the mean soft or ture. And then carry it out! I have to go back to Venice, and tell them about these sad events.	IAGO GRATIANO CASSIO

먼저 전체를 막과 장으로 정리하고 대사 순서대로 나열하여 번호를 매겼다. 대화의 발화자와 청자 그리고 참여한 등장인물, 그리고 대화에서 언급된 인물을 표로 정리하여 나타내었다. 대화에서 짐작할 수 있는 장소와 시간도 구분하였

다. 총 대사 수는 1,210이다. 작품의 크기 별로 분류했을 때 두 번째로 크다. 극 *Othello, the Moor of Venice* 는 청자가 최대 세 명까지 등장한다.

극 *Othello, the Moor of Venice*에 등장하는 모든 등장인물들에 대한 통계 자료를 정리한 표는 <Table 5>와 같다.

<Table 5> General Information of Characters in *Othello, the Moor of Venice*

N	text files	tokens (running words)	types (distinct words)	type/token ratio (TTR)	standardized TTR
	Overall	26,596	3,631	13.65	41.59
1	IAGO	8,634	1,850	21.43	41.99
2	OTHELLO	6,457	1,587	24.58	42.23
3	DESDEMONA	2,824	773	27.37	38.35
4	CASSIO	2,017	696	34.51	41.65
5	EMILIA	1,862	581	31.20	39.70
6	BRABANTIO	1,070	450	42.06	42.90
7	RODERIGO	871	359	41.22	
25	GENTLEMAN	6	6	100.00	

극 *Othello, the Moor of Venice*에는 25명의 인물이 등장한다. 총 25명의 등장인물 가운데 각 인물이 가지는 Token이 1,000이상인 등장인물인 Iago, Othello, Desdemona, Cassio, Emilia, Brabantio만으로 감정 분석과 네트워크 분석을 시행하였다.

### 3.1.1.3 *Antony and Cleopatra*

극 *Antony and Cleopatra*의 대화코퍼스 정보를 나타낸 표는 아래 <Table 6>과 같다.

<Table 6> Dialogue Corpus Information of *Antony and Cleopatra*

N	Act(Scene)	Speaker	Dialogue	Listener(s)
1	1(1)	PHILO	No, our general' s infatuation is out of control. His eyes used to glow with pride when he reviewed his troops. Now his eyes devote themselves exclusively to a certain brown-skinned face. His heart used to burst the buckles on his breastplate in great fights, but now he' s lost all temperance and dedicates his heart to satisfying the lust of an Egyptian whore.	DEMETRIUS
421	2(6)	POMPEY	You should know that when I came here I was prepared to accept your offer. But Mark Antony made me somewhat angry.	CAESAR ANTONY LEPIDUS
...				
1288	5(2)	CAESAR	Come, Dolabella. See that there is a dignified splendor to this great solemnity.	DOLABELLA

위 <Table 6>은 나머지 두 작품들과 마찬가지로 극 전체를 막과 장으로

정리하고 대사 순서대로 나열하여 번호를 매겼다. 대화의 발화자와 청자 그리고 참여한 등장인물, 그리고 대화에서 언급된 인물을 표로 정리하여 나타내었고 대화에서 짐작할 수 있는 장소와 시간도 구분하였다. 총 대사 수는 1,288이다. 작품의 크기 별로 분류했을 때 가장 큰 것을 확인 할 수 있다. 즉 *Antony and Cleopatra*는 청자가 최대 세 명까지 등장한다.

*Antony and Cleopatra*에 등장하는 모든 등장인물들에 대한 통계 자료를 정리한 표는 아래 <Table 7>과 같다.

<Table 7> General Information of Characters in *Antony and Cleopatra*

N	text files	tokens (running words)	types (distinct words)	type/token ratio (TTR)	standardised TTR
	Overall	24,531	3,699	15.08	43.91
1	MARK ANTONY	6,173	1,551	25.13	43.77
2	CLEOPATRA	4,866	1,270	26.10	42.10
3	OCTAVIUS	2,954	970	32.84	45.95
4	ENOBARBUS	2,524	893	35.38	45.90
5	POMPEIUS	943	432	45.81	
	...				
54	TAURUS	2	2	100.00	

즉 *Antony and Cleopatra*에는 54명의 인물이 등장한다. 총 54명의 등장인물 가운데 각 인물이 가지는 Token이 1,000이상인 등장인물인 Mark Antony, Cleopatra, Octavius, Enobarbus만으로 감정 분석과 네트워크 분석을 시행하였다.

### 3.1.2 사랑 희극

본 연구의 범위인 세 편의 희극 작품을 현대어 버전으로 대화코퍼스를 구축하였다. 현대어 버전은 SparkNotes (<http://nfs.sparknotes.com>)에서 가져와 사용하였다. 이번 연구에서 사용된 세 편의 희극 작품을 love comedies라고 분류하고 목록은 *The Two Gentlemen of Verona* (1589), *A Midsummer Night's Dream* (1595), *As You Like It* (1599) 이다. 모든 데이터는 데이터 전처리 과정과 정규화 과정을 거친다.

사랑 희극을 연도별로 정리한 통계 자료는 <Table 8>과 같다. 세 편의 희극을 제작연도 순으로 번호를 매겼다.

<Table 8> General Information of Three Love Comedies

N	Chronology	text files	tokens (running words)	types (distinct words)	type/token ratio (TTR)	standardised TTR
		Overall	55,381	5,791	10.46	40.14
1	1589	The Two Gentlemen of Verona	17,303	2,628	15.19	39.03
2	1595	A Midsummer Night's Dream	16,439	2,901	17.65	42.08
3	1599	As You Like It	21,639	3,167	14.64	39.57

세 편의 작품을 단어 수 별로 정리하면 *As You Like It* > *The Two Gentlemen of Verona* > *A Midsummer Night's Dream* 순이다. Token 수를 기준으로 정리하면 *As You Like It* > *The Two Gentlemen of Verona* > *A Midsummer*

*Night's Dream* 순이다. 다음으로 Type 크기로 정리하면 *As You Like It* > *A Midsummer Night's Dream* > *The Two Gentlemen of Verona* 순이다. 다음으로 STTR값을 기준으로 정리해보면 *A Midsummer Night's Dream* > *The Two Gentlemen of Verona* > *As You Like It* 순이다. 종합해보면 *A Midsummer Night's Dream*은 사용된 단어 수에 비해 새로운 단어가 많이 사용되었다는 것을 알아 낼 수 있었고, *As You Like It*은 사용된 단어 수에 비해 중복된 단어가 많이 사용되었다는 것을 알 수 있다.

### 3.1.2.1 *The Two Gentlemen of Verona*

극 *The Two Gentlemen of Verona*의 대화코퍼스 정보를 나타낸 표는 아래 <Table 9>와 같다.



<Table 9> Dialogue Corpus Information of *The Two Gentlemen of Verona*

N	Act(Scene)	Speaker	Dialogue	Listener(s)
1	1(1)	VALENTINE	Cease to persuade, my loving Proteus: Home-keeping youth have ever homely wits. Were't not affection chains thy tender days To the sweet glances of thy honour'd love, I rather would entreat thy company To see the wonders of the world abroad Than, living dully sluggardiz'd at home, Wear out thy youth with shapeless idleness. But since thou lov'st, love still, and thrive therein, Even as I would when I to love begin.	PROTEUS
...				
483	3(1)	VALENTINE	No.	PROTEUS LAUNCE
...				
867	5(4)	VALENTINE	Please you, I 'll tell you as we pass along, That you will wonder what hath fortunèd.— Come, Proteus, 'tis your penance but to hear The story of your loves discoverèd. That done, our day of marriage shall be yours: One feast, one house, one mutual happiness.	

극 전체를 막과 장으로 정리하고 대사 순서대로 나열하여 번호를 매겼다. 대화의 발화자와 청자 그리고 참여한 등장인물, 그리고 대화에서 언급된 인물을 정리하였고 대화에서 짐작할 수 있는 장소도 정리하였다. 작품에서 청자는 최대 두 명까지 등장한다.

극 *The Two Gentlemen of Verona*에 등장하는 모든 등장인물들에 대한 통계

자료를 정리한 표는 <Table 10>와 같다.

<Table 10> General Information of Characters in *The Two Gentlemen of Verona*

N	text files	tokens (running words)	types (distinct words)	type/token ratio (TTR)	standardised TTR
	Overall	17,303	2,628	15.19	39.64
1	PROTEUS	3,348	968	28.91	41.93
2	VALENTINE	2,888	871	30.16	40.45
3	JULIA	2,500	782	31.28	40.50
4	SPEED	1,469	483	32.88	35.30
5	LAUNCE	1,879	544	28.95	33.40
6	DUKE	1,568	565	36.03	41.30
7	SILVIA	1,169	419	35.84	38.30
8	LUCETTA	570	268	47.02	
18	OUTLAW	6	2	33.33	

극 *The Two Gentlemen of Verona*에는 총 18명의 인물이 등장한다. 희극 작품의 주요 등장인물로 네트워크 분석을 하지 않기 때문에 따로 언급하지 않았다.

### 3.1.1.2 *A Midsummer Night's Dream*

극 *A Midsummer Night's Dream* 의 대화코퍼스의 정보를 나타낸 표는 아래 <Table 11>과 같다.

<Table 11> Dialogue Corpus Information of *A Midsummer Night's Dream*

N	Act(Scene)	Speaker	Dialogue	Listener(s)
1	1(1)	THESEUS	Now, fair Hippolyta, our nuptial hour Draw so napace; four happy days bring in Another moon; but, oh, me thinks, how slow This old moon wanes! She lingers my desires, Like to a step - dame or adowager, Long with eringouta young man's revenue.	HIPPOLYTA
...				
84	1(2)	BOTTOM	I grant you, friends, if you should fright the ladies Out of their wits, they would have no more discretion but to hang us: but I will aggravate my voice so, that I will roar you as gently as any sucking dove; I will roar you an't were any Nightingale.	QUINCE SNUG STARVELING
...				
505	5(2)	PUCK	If we shadows have offended, Think but this,--and all is mended,--That you have but slumber'd here While these visions did appear. And this weak and idle theme, No more yielding but a dream, Gentles, do not reprehend; If you pardon, we will mend. And, as I am an honest Puck, If we have unearned luck Now to 'scape the serpent's tongue, We will make amends ere long; Else the Puck a liar call: So, good night unto you all. Give me your hands, if we	PUCK

		be friends, And Robin shall restore amends.	
--	--	--	--

극 전체를 막과 장으로 정리하고 대사 순서대로 나열하여 번호를 매겼다. 대화의 발화자와 청자 그리고 참여한 등장인물, 그리고 대화에서 언급된 인물을 정리하였고 대화에서 장소에 대한 언급이 없어 빈칸으로 비워두고 정리하였다. 작품에서 청자는 최대 세 명까지 등장한다.

극 *A Midsummer Night's Dream*에 등장하는 모든 등장인물들에 대한 통계 자료를 표로 정리하여 <Table 12>로 나타내었다.

<Table 12> General Information of Characters in *A Midsummer Night's Dream*

N	text files	tokens (running words)	types (distinct words)	type/token ratio (TTR)	standardised TTR
	Overall	16,439	2,901	17.65	42.80
1	THESEUS	1,761	689	39.13	45.50
2	HELENA	1,854	621	33.50	37.90
3	BOTTOM	1,672	540	32.30	38.10
4	OBERON	1,649	652	39.54	44.90
5	LYSANDER	1,420	517	36.41	40.60
6	PUCK	1,415	632	44.66	47.40
7	HERMIA	1,320	512	38.79	43.10
8	DEMETRIUS	1,178	446	37.86	39.90
9	TITANIA	1,189	540	45.42	47.80
10	QUINCE	683	271	39.68	
			...		
31	ALL FOUR	4	4	100.00	

극 *A Midsummer Night's Dream*에는 총 31명의 인물이 등장한다. 희극 작품

에서 등장하는 주요 등장인물들은 네트워크 분석을 하지 않으므로 따로 언급하지 않았다.

### 3.1.2.3 *As You Like It*

극 *As You Like It*의 대화코퍼스의 정보를 나타낸 표는 아래 <Table 13>와 같다.

<Table 13> Dialogue Corpus Information of *As You Like It*

N	Act(Scene)	Speaker	Dialogue	Listener(s)
1	1(1)	ORLANDO	As I remember, Adam, it was upon this fashion bequeathed me by will but poor a thousand crowns, and, as thou sayest, charged my brother on his blessing, to breed me well: and there begins my sadness. My brother Jaques he keeps at school, and report speaks goldenly of his profit: for my part, he keeps me rustically at home, or, to speak more properly, stays me here at home unkept; for call you that keeping for a gentleman of my birth, that differs not from the stalling of an ox? His horses are bred better; for, besides that they are fair with their feeding, they are taught their manage, and to that end riders dearly hired: but I, his brother, gain nothing under him but growth, for the which his animals on his dunghills are as much bound to him as I. Besides this nothing that he so plentifully gives me, the something that nature gave me, his countenance seems to take from me: he lets me feed with his hinds, bars me the place of a brother, and, as much as in him lies, mines my gentility with my	ADAM

			education. This is it, Adam, that grieves me; and the spirit of my father, which I think is within me, begins to mutiny against this servitude. I will no longer endure it, though yet I know no wise remedy how to avoid it.	
--	--	--	---	--

...

96	1(2)	TOUCH STONE	Thus men may grow wiser every day: it is the first time that ever I heard breaking of ribs was sport for ladies.	CELIA ROSALIND LE BEAU
----	------	----------------	--	------------------------------

...

812	5(4)	ROSALIND	It is not the fashion to see the lady the epilogue; but it is no more unhandsome than to see the lord the prologue. If it be true that good wine needs no bush, 'tis true that a good play needs no epilogue; yet to good wine they do use good bushes, and good plays prove the better by the help of good epilogues. What a case am I in then, that am neither a good epilogue, nor cannot insinuate with you in the behalf of a good play! I am not furnished like a beggar, therefore to beg will not become me: my way is, to conjure you; and I'll begin with the women. I charge you, O women! for the love you bear to men, to like as much of this play as please you: and I charge you, O men! for the love you bear to women,—as I perceive by your simpering none of you hate them,—that between you and the women, the play may please. If I were a woman I would kiss as many of you as had beards that pleased me, complexions	ALL
-----	------	----------	---	-----

		that liked me, and breaths that I defied not; and, I am sure, as many as have good beards, or good faces, or sweet breaths, will, for my kind offer, when I make curtsy, bid me farewell.	
--	--	---	--

극 전체를 막과 장으로 정리하고 대사 순서대로 나열하여 번호를 매겼다. 대화의 발화자와 청자 그리고 참여한 등장인물, 그리고 대화에서 언급된 인물을 표로 정리하여 나타내었고 대화에서 짐작할 수 있는 장소도 함께 정리하였다. 작품에서 청자는 최대 세 명까지 등장한다.

극 *As You Like It*에 등장하는 모든 등장인물들에 대한 통계 자료를 표로 정리한 것은 아래 <Table 14>와 같다.

<Table 14> General Information of Characters in *As You Like It*

N	text files	tokens (running words)	types (distinct words)	type/token ratio (TTR)	standardised TTR
	Overall	21,639	3,167	14.64	40.46
1	ROSALIND	5,493	1,267	23.07	39.56
2	ORLANDO	2,399	735	30.64	40.20
3	CELIA	2,158	727	33.69	41.45
5	TOUCHSTONE	2,391	727	30.41	39.10
6	JAQUES	1,966	685	34.84	40.70
7	OLIVER	1,191	471	39.55	42.70
8	DUKES	1,064	454	42.67	43.70
9	PHEBE	715	273	38.18	
24	DENNIS	17	16	94.12	

극 *As You Like It*에는 총 24명의 인물이 등장한다. 그러나 앞서와 마찬가지로 희극 작품의 주요 등장인물로는 네트워크 분석을 하지 않으므로 따로 언급하지 않았다.



## 3.2 방법론

앞에서 언급한 대로, 본 연구에서 사용할 데이터는 셰익스피어의 작품 중 세 편의 비극 작품(*Romeo and Juliet*, *Othello*, *the Moor of Venice*, *Antony and Cleopatra*)과 세 편의 희극 작품(*The Two Gentlemen of Verona*, *A Midsummer Night's Dream*, *As You Like It*)의 텍스트에서 독백, 방백, 노래를 제외한 등장인물끼리의 대화를 기준으로 하여 코퍼스를 수작업으로 구축한다. 원시 자료를 수집 및 가공하고 그 가공된 데이터를 기반으로 감성 분석을 하기 위해 VADER 감성 단어 사전을 활용하여 감성을 분석을 하여 기존의 AFINN을 활용한 결과 값과 비교해 본다. 다음으로 가공된 데이터에 대한 네트워크를 시각화 하기 위해 NetMiner 4.0 이라는 프로그램을 활용했다.

### 3.2.1 감성 분석

본 연구에서는 VADER 감성분석 사전을 활용하여 감성 분석을 하였다. 감성 분석하기에 앞서 먼저 서로 다른 크기의 데이터들을 정규화(Normalization)해야 한다. 데이터를 정규화하는 방법에 대해서 설명한 표는 아래 <Table 15(a), (b)> 와 같다.

<Table 15 (a)> Normalization Example of Three Love Tragedies

	1막	2막	3막	4막	5막
Romeo and Juliet	261	216	223	131	92
Othello, the Moor of Venice	167	197	288	301	257
Antony and Cleopatra	214	247	313	240	174

세 편의 비극 작품에서 각 막 별 대사의 수를 알아낸다. 세 편의 비극 작품 *Romeo and Juliet*, *Othello, the Moor of Venice*, *Antony and Cleopatra*을 각각 A, B, C라고 편의상 나타냈다. 작품을 막(Act)을 기준으로 정규화하려면 세 편의 작품들 중 Act별로 길이가 가장 긴 Act의 대사 수에 나머지 작품들의 대사 길이를 맞춰야 한다.

<Table 15 (b)> Normalization Example of Three Love Tragedies

	1막	2막	3막	4막	5막
A	[1,2,3...259,260,261]	[1,2,14,3,28...244.72, 245.86, 247]	[1,2,40,3,80...310.20, 311.60, 313]	[1,3,30,5,60...296.40, 298.70, 301]	[1,3,79,6,58...251.42, 254.21, 257]
B	[1,2,56,4,12...257.88, 259.44, 261]	[1,2,25,3,5...244.5, 245.75, 247]	[1,2,09,3,18...310.82, 311.91, 313]	[1,2,3...299,300,301]	[1,2,3...255,256,257]
C	[1,2,22,3,44...258.56, 259.78, 261]	[1,2,3...245,246,247]	[1,2,3...311,312,313]	[1,2,25,3,5...298.5, 299.75, 301]	[1,2,48,3,96...254.04, 255.52, 257]

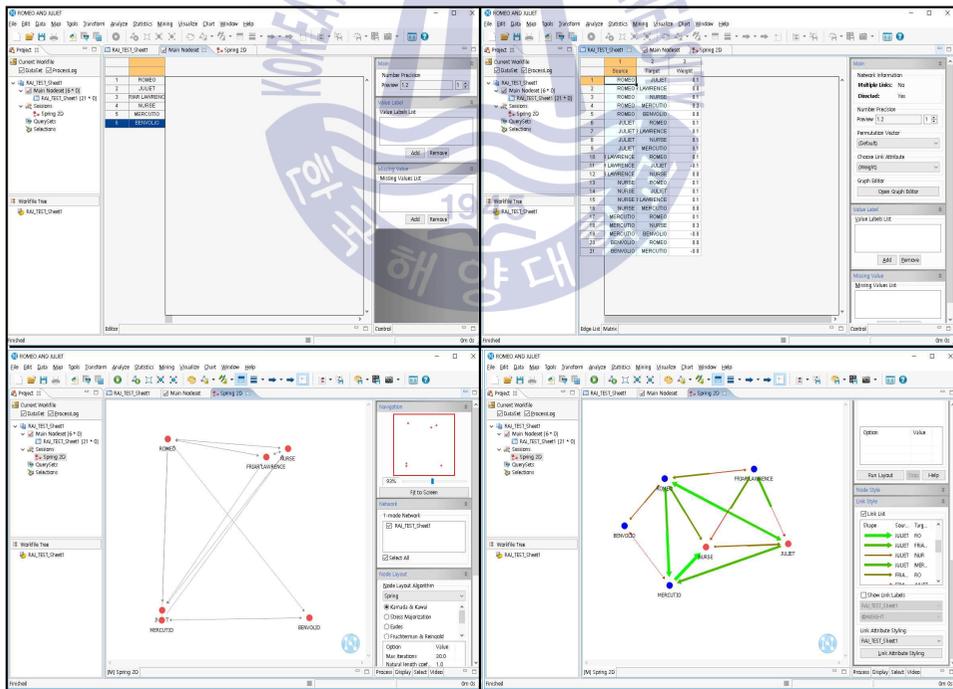
1막에서 가장 긴 작품은 A이므로 B, C의 Act 1 Index사이의 간격을 각각 (A의

Act 1의 길이) / (B의 Act 1의 길이), (A의 Act 1의 길이) / (C의 Act 1의 길이)로 바꿔준다. (A의 Act 1의 길이) / (C의 Act 1의 길이) = 261 / 214 = 1.2196261... ≈ 1.22이다. 그러므로 C의 Act 1의 Index의 간격이 1이 아닌 1.22가 된다. 따라서 바뀐 Index 간격을 적용하면 C는 [1, 2.22, 3.44 . . . . 258.56, 259.78, 261]가 된다. 나머지 B, C의 Act도 같은 방법으로 정규화 한다. 따라서 위의 <Table 15 (a), (b)>에서 보는 것처럼 세 작품의 Act 1의 마지막 대사 번호는 261로 동일하다.



### 3.2.2 네트워크 분석

수많은 프로그램들이 서로 다른 노드 간의 네트워크를 분석하기 위해 개발되고 있다. 여러 네트워크 분석도구들 중 이번 연구에서는 네트워크를 시각화하기 위해 NetMiner 4.0 (Cyram, 2016)을 활용했다. NetMiner는 거대한 양의 데이터에 대한 네트워크를 시각화하여 보여주는 소프트웨어 도구이다. 이는 네트워크의 근본적인 패턴과 구조를 파악하도록 만들어졌으며, 일반적인 연구를 하거나 정보 과학, 생물학, 지리학, 생태학, 정치학과 같은 다양한 분야에 대한 소셜 네트워크 가르치는데 사용되어 왔다. NetMiner 4.0을 사용하여 네트워크를 연결한 단계별 캡처는 아래 <Fig. 11>과 같다.



<Fig. 11> Work Platform of NetMiner 4.0

위의 <Fig. 11>은 극 *Romeo and Juliet*을 실례로 들어 네트워크를 연결시키는 4 단계를 시계방향의 순서로 캡처한 것이다. 먼저 NetMiner 프로그램에 극 *Romeo and Juliet*의 주요 등장인물들의 대화코퍼스를 입력시킨다. 각 등장인물 간의 대화코퍼스가 가지는 점수가 자동으로 추출된다. 추출된 대화코퍼스의 점수를 기반으로 1차적인 네트워크를 연결시킨다. 마지막으로 등장인물 간의 누적 감정 분석 값으로 2차 네트워크를 그려 완성한다. 따라서 마지막 단계의 네트워크에서는 등장인물 간 연결 정도가 어느 강도로 연결되어있는지를 파악할 수 있다.

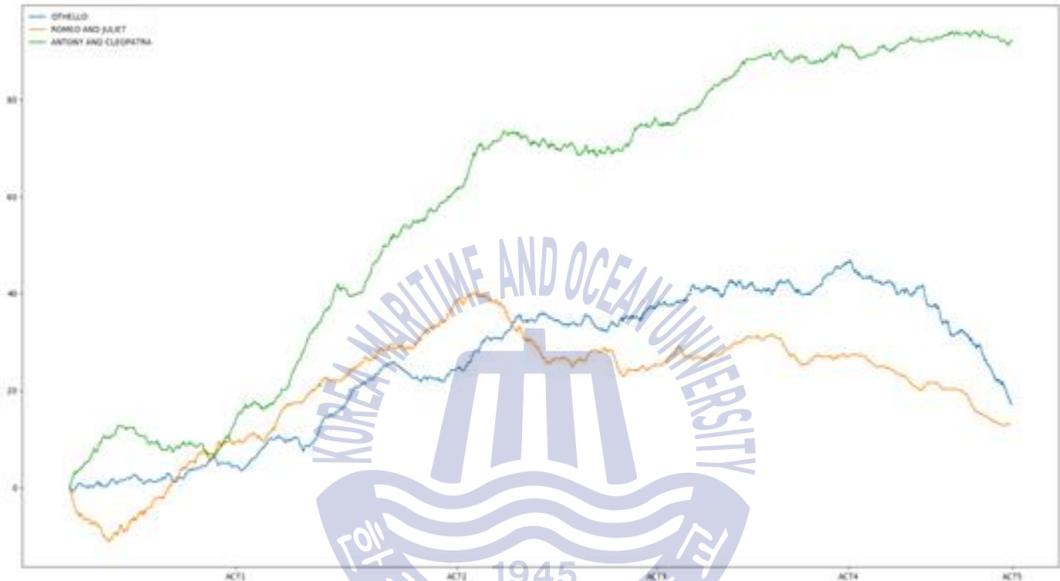


## 제 4 장 사랑 비극의 감정 분석

4.1에서는 세 편의 사랑 비극의 전체 대사를 VADER 감정 사전을 활용하여 감정 분석하여 그 값들의 누적 VADER 감정값으로 그래프를 그렸다. 세 편의 사랑 비극이 보여주는 그래프를 통해 비극의 특징을 나타낼 수 있는 세 가지 패턴을 찾아보고자 한다. 첫째, 모든 등장인물들의 대화코퍼스에서 누적된 VADER 감정값의 패턴을 찾아본다. 둘째, 각 그래프에서 가장 높은 누적 VADER 감정값을 가지는 대화의 패턴을 찾아본다. 셋째, ‘love’가 포함된 대화의 패턴을 찾아본다. 4.2에서는 4.1에서 살펴본 첫 번째 패턴의 신뢰성을 확보하기 위하여 세 편의 사랑 희극의 전체 대사를 VADER 감정 사전을 활용하여 감정 분석하여 그 값들의 누적 VADER 감정값으로 그래프를 그려 비교하였다. 마찬가지로 그래프를 통해 희극의 특징을 나타내는 패턴을 찾아보고자 한다. 4.3에서는 세편의 사랑 비극과 세편의 사랑 희극의 전체 대사를 VADER 감정 사전을 활용하여 감정 분석하여 그 값들의 누적 VADER 감정값을 하나의 그래프로 나타냈다. 여섯 편의 작품을 하나의 그래프로 비교하여 나타나는 특징들을 설명 하고자 하였다. 또한 작품 별 누적 값이 가장 높은 점의 대사를 찾아본다. 4.4에서는 세 편의 사랑 비극에 등장하는 남성 주인공과 여성 주인공의 서로의 대사를 감정 분석하여 VADER 점수와 AFINN 점수의 누적 값으로 그래프를 그려 비교하고, 작품 별 주요 등장인물들의 대사를 감정 분석하여 막별로 구분하고 그 값들에 대해 비교한다. 마지막 4.5에서는 4장을 요약하였다.

#### 4.1 사랑 비극 비교

셰익스피어 세 편의 비극 작품의 VADER 감정 사전을 활용하여 막별 감정 분석한 그래프는 아래 <Fig. 12>와 같다.



<Fig. 12> Sentiment Analysis of Three Love Tragedies

위 <Fig. 12>은 세 편의 비극 작품의 대사 Index를 언급한 방법대로 정규화 하여 1막부터 5막까지 동일한 길이로 처리하였다. 그리고 대화코퍼스의 대사에 해당하는 VADER 값을 누적하여 그래프를 그렸다. 위에서부터 *Antony and Cleopatra*, *Romeo and Juliet*, *Othello, the Moor of Venice* 순으로 그래프가 그려진다. 극의 후반부인 5막에서 상승하는 그래프를 보여주는 유형과 5막에서 하강하는 그래프를 보여주는 유형, 두 가지로 나누어진다. 첫 번째 유형에 해당하는 *Antony and Cleopatra*에서 2막까지는 계속 긍정적인 대화가 이루어지다가 3막에서 부정적인 대화를 가진다. 다시 4막부터 5막까지 완만한 상승 곡선을 갖는다. 두 번째 유형에 해당하는 *Romeo and Juliet*, *Othello, the Moor of*

*Venice*는 유사한 패턴을 보이거나 *Romeo and Juliet*은 2막까지는 긍정적인 대화가 많고 3막부터 3막의 중간 부분까지는 부정적인 대화가 많아진다. 그 이후부터는 중간적인 입장을 유지하는데 극의 마지막인 5막에 완만한 하강의 곡선을 갖는다. 반면에 *Othello, the Moor of Venice*는 4막까지는 긍정적인 대화가 점진적으로 이루어지다가 극의 후반부인 5막부터는 급격한 하강 곡선을 보여준다.

사랑 비극 세 편의 그래프들을 비교해보면 모든 그래프들이 1막 마지막 부분을 기준으로 교차됨과 동시에 다양한 변화를 보이고 있다. 이는 1막의 마지막 부분이 세 편의 비극 작품의 터닝 포인트로 작용한다는 것을 알 수 있다.

그렇다면 <Fig. 12>에서 보이는 각 작품의 그래프에서 가장 높은 누적 VADER값을 나타내는 부분에서 나타내는 사랑 비극의 패턴을 찾을 수 있는가를 살펴본다.

극 *Romeo and Juliet* 전체 대화코퍼스에 대해 가장 높은 VADER 누적 값을 가진 대사와 그 앞 뒤 대사의 VADER 누적 값을 나타낸 표는 아래 <Table 16>와 같다.

<Table 16> Cumulative VADER Valences of *Romeo and Juliet*

N	Act (Scene)	Speaker	Dialogue	Listener	VADER	Cumulative VADER
497	3(1)	TYBALT	Well, may peace be with you. Here comes my man, the man I'm looking for.	MERCUTIO	0.3404	40.1395
498	3(1)	MERCUTIO	He's not your man. Alright, walk out into a field, and he'll chase you. In that sense you can call him your "man."	TYBALT	0.0833	40.2228
499	3(1)	TYBALT	Romeo, there's only one thing I can call you. You're a villain.	ROMEO	-0.2787	39.9441

위 <Table 16>에서 알 수 있듯이 VADER 누적 값이 가장 높은 대사는 대사번호 498번으로 Mercutio가 Tybalt에게 한 대사로 40.2228로 가장 높았고 해당 대사에서는 ‘love’ 라는 단어는 찾아볼 수 없었다.

극 *Othello, the Moor of Venice* 전체 대화코퍼스에 대해 VADER 누적 값이 가장 높은 대사와 그 앞 뒤 대사의 VADER 값을 나타낸 표는 아래 <Table 17>과 같다.



<Table 17> Cumulative VADER Valences of *Othello, the Moor of Venice*

N	Act (Scene)	Speaker	Dialogue	Listener	VADER	Cumulative VADER
956	5(1)	IAGO	I' m right behind you. Be bold, and get ready.	RODERIGO	0.3125	46.6990
957	5(1)	RODERIGO	I don' t really want to do this, but he' s given me good reasons. I guess it' s only one man—no big deal. My sword comes out, and he dies.		0.2102	46.9092
958	5(1)	IAGO	(to himself) I' ve rubbed this young pimple until he' s ready to pop, and now he' s angry. Whether he kills Cassio, or Cassio kills him, or they kill each other, it all works in my favor. If Roderigo survives, though, he' ll ask me for all the gold and jewelry that I stole from him and said I gave to Desdemona. I can' t let that happen. If Cassio survives, he' s so handsome and well-spoken that he makes me look ugly. And besides, the Moor might tell him about my lies about him.—That would be very dangerous for me. No, he' s got to die. Let it be so. I hear him coming.		-0.2592	46.6500

위 <Table 17>에서 알 수 있듯이 VADER 누적 값이 가장 높은 대사는 대사번호 957번으로 Roderigo가 한 대사로 46.9092로 가장 높았고 해당 대사에서는

‘love’ 라는 단어는 찾아볼 수 없었다.

극 *Antony and Cleopatra*의 VADER 누적 값이 가장 높은 대사와 그 앞 뒤 대사들의 VADER 누적 값을 나타낸 표는 아래 <Table 18>과 같다.

<Table 18> Cumulative VADER Valences of *Antony and Cleopatra*

N	Act (Scene)	Speaker	Dialogue	Listener	VADER	Cumulative VADER
1248	5(2)	A COUNTRY MAN	You must know that the snake will act according to his nature.	CLEOPATRA	0.5267	93.7767
1249	5(2)	CLEOPATRA	Yes, yes. Farewell.	A COUNTRY MAN	0.3298	94.1065
1250	5(2)	A COUNTRY MAN	Listen, the snake is dangerous unless handled by people who are used to him. There isn't any kindness in the snake.	CLEOPATRA	-0.009	94.0975

위 <Table 18>에서 알 수 있듯이 대사 번호 1,250번의 VADER 누적 값이 94.1065로 가장 높았고 해당 대사에서는 ‘love’ 라는 단어는 찾아볼 수 없었다.

세 편의 비극 작품의 가장 높은 누적 VADER값을 나타내는 대사에서 ‘love’ 라는 단어를 발견하지 못하였다. 따라서 누적 VADER 감정 점수는 사랑 비극의 패턴을 보여주지 못한다.

지금까지 선그래프에서 발견한 것은 다음의 두 가지이다. 첫째, 뚜렷하게 관찰된 것은 5막에서 누적 감정값이 두 가지 유형으로 나뉘어 급격하게 하향하거나 가장 마지막 부분에서 천천히 하향하는 감정 패턴이다. 둘째, 선그래프의 가

장 높은 누적 값을 보이는 감정 패턴에서는 사랑 비극의 뚜렷한 특징을 보여주지 못했다.

세 번째 감정 패턴을 고려할 수 있는 것은 사랑 비극에서 ‘love’가 포함된 작품 별 대화코퍼스를 구축하여 이러한 코퍼스에서 어떤 감정의 패턴을 찾을 수 있는지를 살펴보는 것이다. 우선 *Romeo and Juliet*의 ‘love’가 포함된 대화 코퍼스에서 대사의 상위 VADER값으로 정렬한 기본 통계는 아래 <Table 19>과 같다.

<Table 19> Dispersion of Dialogues including ‘Love’ in *Romeo and Juliet*

VADER	1막	2막	3막	4막	5막	Total
0.9XXX		2				2
0.8XXX	1					1
0.7XXX						0
...						
-0.2XXX	1	1	1	2	1	6
-0.3XXX			1	1		2
-0.4XXX			1			1

위 <Table 19>은 ‘love’가 포함된 *Romeo and Juliet* 대사의 VADER 값 분포도 VADER 점수의 분포를 표로 나타낸 것이다. 2막에서 가장 높은 VADER점수대인 0.9가 나타났다. 0.7의 점수대는 나타나지 않았다. 반대로 가장 낮은 점수대인 -0.4는 3막에서 나타났다. -0.2의 점수대는 극의 전반에 분포되어있는 것을 확인할 수 있었다.

한편 VADER값이 가장 높은 점수와 가장 낮은 VADER점수를 보여주는 실례를 살펴보자. ‘Love’가 포함된 대화코퍼스의 상위 VADER의 고득점 값을 갖는 구체적인 대사로 정리한 표는 아래 <Table 20>과 같다.

<Table 20> Dialogues including ‘Love’ in *Romeo and Juliet*

N	Act(Scene)	Speaker	Dialogue	Listener	VADER
1	2(5)	NURSE	Your love says, like an honorable gentleman, who is courteous, kind, handsome, and, I believe, virtuous— where is your mother?	JULIET	0.9643
2	2(2)	ROMEO	I would be satisfied if we made each other true promises of love.	JULIET	0.9081
3	4(1)	PARIS	You will also confess, I’ m sure, that you love me.	JULIET	0.7579
...					
109	3(5)	LADY CAPULET	Will you cry about your cousin’ s death forever? Are you trying to wash him out of his grave with tears? If you could, you couldn’ t bring him back to life. So stop crying. A little bit of grief shows a lot of love. But too much grief makes you look stupid.	JULIET	-0.4158

위 <Table 20>은 ‘love’ 가 포함된 대사 109개 중 VADER값이 높은 순서로 나열하였다. VADER값이 가장 높은 대사는 2막 5장에서 Nurse가 Juliet에게 한 대사로 0.9643의 점수를 가진다. 반대로 3막 5장에서 Lady Capulet이 Juliet에게 한 대사로 -0.4158으로 가장 낮았다.

다음으로 극 *Othello, the Moor of Venice*의 전체 대화코퍼스에서 ‘love’ 를 포함한 대사의 통계는 아래 <Table 21>과 같다.

<Table 21> Dispersion of Dialogues including ‘Love’ in *Othello, the Moor of Venice*

VADER	1막	2막	3막	4막	5막	Total
0.9XXX		1				1
0.8XXX	1	1				2
0.7XXX						0
...						
-0.3XXX	1		1			2
-0.4XXX					1	1
-0.5XXX			1			1

위 <Table 21>은 VADER 점수에 대한 통계자료를 표로 나타낸 것으로 각 점수 대별 대사의 개수를 나타내었다. 가장 높은 VADER 점수대를 의미하는 0.9는 2막에서 나타났다. 0.7의 점수대는 나타나지 않았다. 반대로 가장 낮은 점수대인 -0.5는 3막에서 나타났다.

극 *Othello, the Moor of Venice*에서 ‘love’가 포함된 대화코퍼스가 가지는 VADER 점수를 구체적인 대사와 함께 정리하여 표로 나타냈으며 아래 <Table 22>과 같다.

<Table 22> Dialogue including ‘Love’ in *Othello, the Moor of Venice*

N	Act (Scene)	Speaker	Dialogue	Listener(s)	VADER
1	2(1)	DESDEMONA	God willing, our love and our happiness will only increase as we get older.	OTHELLO	0.9042
2	1(3)	OTHELLO	In the meantime I’ ll tell you all, as honestly as I confess my	DUKE SENATORS	0.8984

			sins to God, how I wooed this beautiful lady, and how she came to love me.		
3	2(1)	IAGO	A woman who was beautiful but never proud, who could speak well but knew when to be quiet, who dressed well but was never overdressed, who had self-restraint even when she could get what she wanted, a woman who never took revenge, who overlooked it when people hurt her, who was too wise to do anything stupid, who could think without revealing her thoughts, and who could refrain from flirting with men in love with her, that kind of woman, if she ever existed, would—DESDEMONA	DESDEMONA	0.8715
...					
72	3(4)	CASSIO	It's not because I don't love you.	BIANCA	-0.5216

위 <Table 22>는 *Othello, the Moor of Venice*에서 'love' 을 포함 한 대사는 총 72개이다. 그 중 VADER값이 가장 높은 대사는 대사번호 1번으로 2막 1장에서 Desdemona가 Othello에게 한 대사로 VADER값이 0.9042로 가장 높았다. 반대로 3막 4장에서 Cassio가 Bianca에게 한 대사로 -0.5216으로 가장 낮았다.

극 *Antony and Cleopatra*의 전체 대화코퍼스에서 'love' 를 포함한 대사의 통계는 아래 <Table 23>과 같다.

<Table 23> Dispersion of Dialogue including 'Love' in *Antony and Cleopatra*

VADER	1막	2막	3막	4막	5막	Total
0.8XXX	3					3
0.7XXX		2				2
0.6XXX	4	1				5
...						
-0.0XXX	2		3			5
-0.1XXX	1					1
-0.2XXX			2			2

위 <Table 23>은 VADER 감정값에 대한 통계 자료를 표로 나타낸 것으로 각 점수대별 대사의 개수를 나타내었다. 가장 높은 VADER점수대를 의미하는 0.8은 1막에서 3번 나타났다. 0.6의 점수대는 1막에서 4번 나타났고 2막에서 1번 나타나는 분포를 보였다. 반대로 가장 낮은 점수대를 의미하는 -0.2는 3막에서 2번 나타났다. 또한 'love'가 포함된 대화코퍼스가 가지는 점수를 구체적인 대사와 함께 정리하여 표로 나타냈으며 아래 <Table 24>과 같다.

<Table 24> Dialogues including 'Love' in *Antony and Cleopatra*

N	Act (Scene)	Speaker	Dialogue	Listener	VADER
1	1(3)	CHARMIAN	Madam, I think if you love him so much, you aren't using the best way to get him to reciprocate.	CLEOPATRA	0.8555
2	1(2)	THE FORTUNETELLER	You will love more than you are loved.	CHARMIAN	0.8442
3	1(2)	CHARMIAN	I would rather get passion from drink than from love.	THE FORTUNE TELLER	0.8020

...					
53	3(13)	THIDIAS	He knows that you didn' t embrace Antony because you loved him, but because you feared him.	CLEOPATRA	-0.2960

위 <Table 24>는 극 *Antony and Cleopatra*에서 'love' 을 포함 한 대사는 총 53개이다. 그 중 VADER값이 가장 높은 대사는 대사번호 1번으로 1막 3장에서 Charmian이 Cleopatra에게 한 대사로 VADER값이 0.8555로 가장 높았다. 반대로 3막 13장에서 Thidias가 Cleopatra에게 한 대사로 -0.2960으로 가장 낮았다.

그렇다면 각 작품 별 VADER 누적 값이 가장 높은 대사에서 'love' 를 포함 하는 대사가 누적 VADER값도 높은지를 살펴보기 위해 각 작품 별 VADER 누적 값의 순위를 매겨 <Table 25>에서 보여준다.

<Table 25> Rank of Cumulative VADER Valence of *Romeo and Juliet*

순위	N	Act (Scene)	Dialogue	VADER	Cumulative VADER
1	498	3(1)	He' s not your man. Alright, walk out into a field, and he' ll chase you. In that sense you can call him your "man."	0.0833	40.2228
2	497	3(1)	Well, may peace be with you. Here comes my man, the man I' m looking for.	0.3404	40.1395
3	499	3(1)	Romeo, there' s only one thing I can call you. You' re a villain.	-0.2787	39.9441
4	502	3(1)	I disagree. I' ve never done you harm. I <i>love</i> you more than you can understand until you know the reason why I <i>love</i> you. And so,	0.4545	39.8915

			good Capulet—which is a name I love like my own name—you should be satisfied with what I say.		
5	491	3(1)	You’ ll find me ready enough to do that, sir, if you give me a reason.	0.3612	39.8604
6	496	3(1)	Men’ s eyes were made to see things, so let them watch. I won’ t move to please anybody.	0.1591	39.7991
7	500	3(1)	Tybalt, I have a reason to love you that lets me put aside the rage I should feel and excuse that insult. I am no villain. So, goodbye. I can tell that you don’ t know who I am.	-0.2601	39.6840
8	503	3(1)	This calm submission is dishonorable and vile. The thrust of a sword will end this surrender.	-0.2107	39.6808
9	508	3(1)	Noble Mercutio, put your sword away.	0.4588	39.6771
9	509	3(1)	(to TYBALT) Come on, sir, perform your forward thrust, your passado.	0	39.6771
...					
923	63	1(1)	He’ s been seen there many mornings, crying tears that add drops to the morning dew and making a cloudy day cloudier with his sighs. But as soon as the sun rises in the east, my sad son comes home to escape the light. He locks himself up alone in his bedroom, shuts his windows to keep out the beautiful daylight, and makes himself an artificial night. This mood of his is going to bring bad news, unless someone smart can fix what’ s bothering him.	-0.2597	-11.1177

위 <Table 25>에서 극 *Romeo and Juliet*의 누적 VADER값에 대한 상위 10개의 대사에서 ‘love’가 포함되어 있는 대사가 상위에 있는지를 살펴본다. VADER 누적 값은 40.2228로 가장 높았고 가장 낮은 VADER 누적 값은 -11.1177이다. ‘love’가 포함된 대사는 4위로 39.8915의 누적 값을 가진다. 그러나 해당 대사의 VADER 감정 점수와 누적 VADER 감정 점수가 사랑 비극의 특징을 찾는 패턴을 찾는 데 의미를 가지지 못한다. 나머지 상위 10개에서는 ‘love’가 포함된 대사는 찾아볼 수 없었다.

다음 *Othello, the Moor of Venice*에서 VADER 누적 값의 순위를 알아보고 상위 10위에서 ‘love’가 포함되어 있는 대사가 있는지 아래 <Table 26>에서 살펴본다.

<Table 26> Rank of Cumulative VADER Valences of *Othello, the Moor of Venice*

Rank	N	Act (Scene)	Dialogue	VADER	Cumulative VADER
1	957	5(1)	I don't really want to do this, but he's given me good reasons. I guess it's only one man—no big deal. My sword comes out, and he dies.	0.2102	46.9092
2	956	5(1)	I'm right behind you. Be bold, and get ready.	0.3125	46.6990
3	958	5(1)	(to himself) I've rubbed this young pimple until he's ready to pop, and now he's angry. Whether he kills Cassio, or Cassio kills him, or they kill each other, it all works in my favor. If Roderigo survives, though, he'll ask me for all the gold and jewelry that I stole from him and said I gave to Desdemona. I can't let that happen. If Cassio survives, he's so handsome and well-spoken that he makes me look ugly.	-0.2592	46.6500

			And besides, the Moor might tell him about my lies about him.—That would be very dangerous for me. No, he's got to die. Let it be so. I hear him coming.		
4	953	4(3)	Good NIGHT, good NIGHT. I pray that God will let me learn from women like that—not to follow their bad example, but to avoid it!	0.1811	46.5671
4	954	5(1)	Here, stand behind this wall; he'll come right away. Keep your sword out, and then stick it in as far as it'll go. Quick, quick. Don't be afraid. I'll be right next to you. This will either make us or break us. Keep that in mind, and be steady.	0	46.5671
6	955	5(1)	Stay right near me. I may mess it up.	-0.1806	46.3865
7	952	4(3)	Yes, a dozen of them—as many as there are women in the world, in fact. But I do think it's the husband's fault if we wives cheat on them. For instance, our husbands may stop sleeping with us, and give it out to other women instead. Or they may get insanely jealous, and keep us from going anywhere. Or let's say they hit us, or cut back on the money they give us out of spite. We have feelings. We may be able to forgive them, but we want to get back at them too. Husbands need to know that their wives are human beings too. They see, smell, and taste sweet and sour just like their husbands. Why do they replace us with other women? Do they do it for fun? I think they do. Is it out of lust? I think so. Is it a weakness? It is. And don't we have passions, and a taste for fun, and weaknesses, just like men? Then	0.0051	46.3860

			tell them to treat us well. Or let them figure out that the bad things we do are just what we learned from them.		
8	951	4(3)	I don' t think any woman like that exists.	0.3612	46.3809
9	949	4(3)	I' d never do such a bad thing, not for the whole world!	0.484	46.3597
10	959	5(1)	I know the sound of his walk.—Die, villain!	-0.5983	46.0517
...					
1210	4	1(1)	I do hate him, I swear. Three of Venice's most important noblemen took their hats off to him and asked him humbly to make me his lieutenant, the second in command. And I know my own worth well enough to know I deserve that position. But he wants to have things his own way, so he sidesteps the issue with a lot of military talk and refuses their request. "I' ve already chosen my lieutenant," he says. And who does he choose? A guy who knows more about numbers than fighting! This guy from Florence named Michael Cassio. He has a pretty wife but he can' t even control her. And he' s definitely never commanded men in battle. He' s got no more hands-on knowledge of warfare than an old woman—unless you count what he' s read in books, which any peace-lover can do. His military understanding is all theory, no practice. But Cassio' s been chosen over me. My career is cut short by some bookkeeper, even though the general saw my fighting skills first-hand in Rhodes and Cyprus. This accountant is now lieutenant, while I end up as the Moor' s flag bearer.	-0.1094	-0.8982

위 <Table 26>을 통해 극 *Othello, the Moor of Venice*의 누적 VADER값에 대한 상위 10개의 대사에서 ‘love’가 포함되어 있는 대사가 상위에 있는지를 살펴본다. VADER 누적 값은 46.9092로 가장 높았고 가장 낮은 VADER 누적 값은 -0.8982이다. 그러나 ‘love’가 포함된 대사는 찾아볼 수 없었다.

극 *Antony and Cleopatra*에서 VADER 누적 값의 순위를 알아보고 상위 10개에서 ‘love’가 포함되어 있는 대사가 있는지 아래 <Table 27>에서 살펴본다.

<Table 27> Rank of Cumulative VADER Valences of *Antony and Cleopatra*

Rank	N	Act (Scene)	Dialogue	VADER	Cumulative VADER
1	1249	5(2)	Yes, yes. Farewell.	0.3298	94.1065
2	1250	5(2)	Listen, the snake is dangerous unless handled by people who are used to him. There isn't any kindness in the snake.	-0.009	94.0975
3	1258	5(2)	Clouds, dissolve into rain, so that I could say the gods themselves are weeping!	-0.4926	94.0791
4	1231	5(2)	I'll never see it. I know my nails are stronger than my eyes; I'll scratch them out.	0.1909	94.0408
5	1225	5(2)	Farewell, and thanks.	0.4404	94.0277
6	1256	5(2)	Yes, truly, I wish you good luck with the snake.	0.9217	93.9507
7	1241	5(2)	Do you have in there the pretty snake of the Nile that kills without pain?	0.3404	93.917
8	1230	5(2)	No, that's the truth.	0.0258	93.8499
9	1229	5(2)	Oh, the good gods!	0.4926	93.8241
10	1222	5(2)	Madam, according to my	0.4747	93.8017

			<p>promise—which my <i>love</i> to you has made a religious vow—I tell you that Caesar intends to travel through Syria. Within three days you and your children will be sent ahead.</p> <p>Make the best use of this information you can. I have done your bidding and fulfilled my promise.</p>		
...					
1288	1	1(1)	<p>No, our general' s infatuation is out of control. His eyes used to glow with pride when he reviewed his troops. Now his eyes devote themselves exclusively to a certain brown-skinned face. His heart used to burst the buckles on his breastplate in great fights, but now he' s lost all temperance and dedicates his heart to satisfying the lust of an Egyptian whore.</p>	0.0231	0.0231

위 <Table 27>를 통해 극 *Antony and Cleopatra*의 누적 VADER값에 대한 상위 10개의 대사에서 ‘love’ 가 포함되어 있는 대사가 상위에 있는지를 살펴본다. VADER 누적 값은 94.1065로 가장 높았고 가장 낮은 VADER 누적 값은 0.0231이다. 그리고 ‘love’ 가 포함된 대사는 10위의 대사에서 찾아볼 수 있었다.

위의 세 개의 표에서 ‘love’ 는 비극 작품 세 편의 주제이기 때문에 자주 등장할 것으로 예상했지만 누적 VADER 값이 가장 높은 대사에서는 발견 할 수 없었다.

## 4.2 사랑 희극 비교

셰익스피어의 세 편의 희극 작품에서 전체 대화코퍼스를 기준으로 VADER 감정 사전을 활용하여 막별로 감정 분석한 그래프는 아래 <Fig. 11>와 같다.



<Fig. 13> Sentiment Analysis of Three Love Comedies

위 <Fig. 13>은 세 편의 희극 작품의 대사 Index를 3장에서 언급한 방법대로 정규화 하여 1막부터 5막까지 동일한 길이로 처리하였다. 그리고 대화코퍼스의 대사에 해당하는 VADER 값을 누적하여 그래프를 그렸다.

극 *As You Like It* 은 1막에서 긍정적인 감정 그래프가 점차 증가한다. 2막에서 3막 초반까지는 긍정적인 감정이 유지되고 있다. 그러나 3막의 중반부에서 급격하게 상승하는 것을 확인 할 수 있으며 그 이후 4막, 5막에서도 지속적인 증가를 보이고 있다. 극 *As You Like It*의 감정 그래프에서 2막 중반 그리고 3막 초반에서 두 개의 작은 피크들을 발견 할 수 있다. 극 *The Two Gentlemen of Verona*에서도 긍정적인 감정이 계속해서 증가하는 모습을 보인다. 극 *The Two Gentlemen of Verona*의 감정 그래프에서도 작은 피크들을 발견 할 수 있

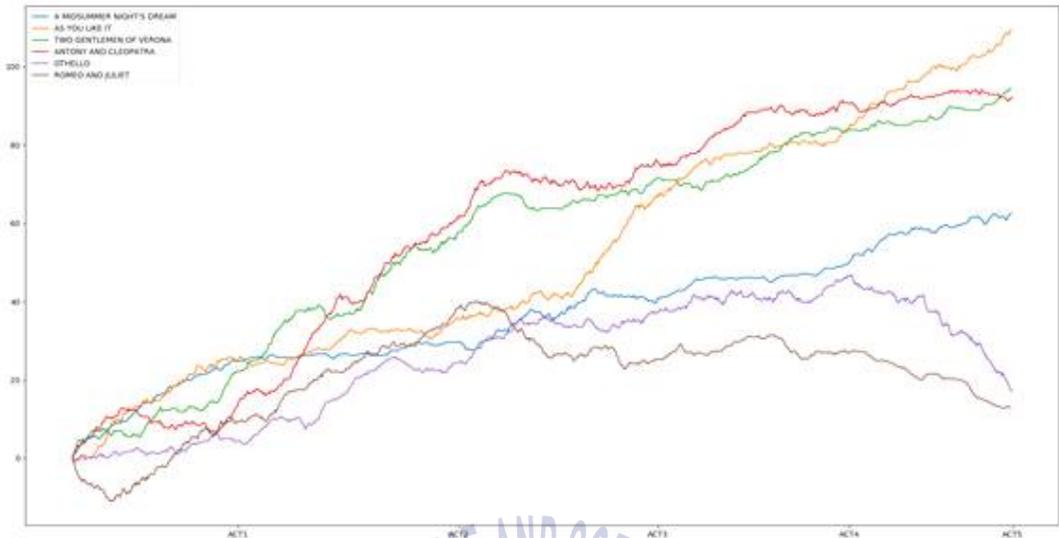
는데 1막 중반에서 하나, 2막 초반에서 하나, 3막 초반에서 하나, 총 세 개의 작은 피크들을 발견 하였다. 극 *A Midsummer Night's Dream*은 1막에서 3막 후반까지 *As You Like It* 과 동일한 상승곡선을 그린다. 그러나 극 *As You Like It* 이 3막 후반에서 급격한 증가를 보이는 반면 극 *A Midsummer Night's Dream*에서는 일정하게 증가하는 모습을 보여준다.

사랑 희극 세 편의 그래프들을 비교해보면 모든 그래프들이 2막 시작부분을 기준으로 교차됨과 동시에 다양한 변화를 보이고 있다. 이는 2막의 초입부가 세 편의 희극 작품의 터닝 포인트로 작용한다는 것을 알 수 있다. 그리고 세 편의 사랑 희극은 가장 높은 누적 VADER 감정 점수를 가지는 피크가 모두 5막에 있다는 것을 관찰하였다.

#### 4.3 사랑 비극과 사랑 희극의 비교

셰익스피어의 작품은 전체 37편으로 비극은 10편, 희극은 17편, 역사극은 20편이다. 앞선 소절에서 ‘사랑’을 주제로 한 비극 작품의 두드러진 특징은 공통적으로 5막에서 누적의 감정값이 급격한 하강 또는 완만한 하강을 선그래프의 시간시각화로 보여주었다. 이러한 사랑 비극의 공통적인 특징이 비극에서만 나타나는 특이한 특징인지를 확인하기 위하여 ‘사랑’을 주제로 한 희극 작품 세 편을 같이 비교해보자.

Archer et al (2005)은 세 편의 비극 작품과 세 편의 희극 작품이라고 분류하였다. 이런 분류를 근거로 사랑의 주제 희비극 총 여섯 편의 작품을 앞에서 보여준 선그래프와 같은 방법으로 막의 흐름에 따라 보여준 누적된 감정의 선그래프는 아래 <Fig. 14>와 같다.



<Fig. 14> Sentiment Analysis of and Three Love Tragedies and Three Love Comedies

위 <Fig. 14>은 세 편의 비극 작품 그리고 세 편의 희극 작품의 모든 등장인물들의 대화코퍼스를 3장에서 언급한 방법대로 정규화 시켜 1막부터 5막까지 동일한 길이로 처리하여 여섯 작품 모두의 감정 변화를 하나의 그래프로 나타냈다. 선그래프를 통해 비극의 특징을 확인하고 희극과 구분 지을 수 있는 특징이 있는지를 관찰하여 흥미로운 점을 발견하였다. 첫 번째 흥미로운 점은 극의 결론 부분인 5막에서 각 작품의 감정 선이 긍정으로 향하는지 (상향식) 부정으로 향하는지 (하향식)에 따라 두 가지 유형으로 나뉜다는 것이다. 마지막 5막이 긍정으로 향하는 유형 (상향식)에는 두 편의 희극 작품 *As You Like It*, *The Two Gentlemen of Verona*와 한 편의 비극 작품 *Antony and Cleopatra*가 포함되어 있다. 그리고 5막이 부정으로 향하는 유형 (하향식)에는 두 편의 비극 작품 *Romeo and Juliet*, *Othello, the Moor of Venice*과 한 편의 희극 *A Midsummer Night's Dream*이 포함된다. 두 번째 흥미로운 점은 희극 작품 *A Midsummer Night's Dream*과 비극 작품 *Antony and Cleopatra*에서 특이한 점을 발견할 수 있다는 것이다. 극 *A Midsummer Night's Dream*은 희극으로 분류되지만 비극 *Othello, the Moor of Venice*와 같이 1막부터 4막까지 비슷한 감

정 그래프를 보여준다. 그러나 여기서 극 *A Midsummer Night's Dream*이 희극으로 분류될 수 있는 이유는 마지막 5막에서 감정 그래프가 상승하는 상향식 유형에 속하기 때문이다. 그리고 극 *Antony and Cleopatra*는 비극으로 분류되지만 1막에서 5막까지 긍정적인 감정값을 나타내는 희극 *The Two Gentlemen of Verona*와 같은 모양으로 완만한 그래프를 보여주고 있다. 그러나 여기서 극 *Antony and Cleopatra*가 비극으로 분류될 수 있는 이유는 5막의 마지막 부분이 하강하는 하향식 유형에 속하기 때문이다.

이러한 누적 감정 점수의 선그래프를 통해 극의 흐름에서 마지막 5막에서 하강하는 비극과 마지막 5막에서 하강하는 희극의 특징을 관찰하였고 이를 통해 비극과 희극의 특징인 5막에서의 분명한 차이를 살펴보고 비극과 희극을 구분 지을 수 있는 단서 또는 패턴을 발견하였다.

#### 4.4 사랑 비극의 주요 등장인물 비교

지금까지는 막의 흐름에 따른 모든 등장인물의 대화코퍼스를 통한 사랑 비극의 작품 간의 감정 분석을 하였다. 지금부터는 사랑 비극의 주요 등장인물 간의 감정 분석을 한다. 우선 본 연구에서 사용한 VADER 감정 사전이 기존의 AFINN 감정 사전보다 우월하다는 사실을 남녀주인공의 감정 추이의 차이점으로 살펴본다.

작품 *Romeo and Juliet*, *Othello*, *the Moor of Venice*, *Antony and Cleopatra*에 등장하는 등장인물들 가운데 대표적인 등장인물인 남자 주인공과 여자 주인공 서로의 대사만을 VADER 감정 사전과 기존의 감정 사전인 AFINN으로 해당 대사내의 단어 별 점수를 매겨 살펴본다. 따라서 4.4 소절에서는 어떤 감정 사전이 주인공의 감정 분석에 더 적합한지를 파악해 본다.

#### 4.4.1 남성 주인공과 여성 주인공 감정 분석

4.4.1.1에서는 극 *Romeo and Juliet*의 남성 주인공인 Romeo 와 여성 주인공인 Juliet 사이의 대화코퍼스를 기준으로 VADER 감정 사전을 활용하여 각 대사에 해당하는 단어들의 값을 알아내고 기존의 감정 사전인 AFINN을 활용한 단어 점수들과도 비교해 본다. 4.4.1.2에서는 극 *Othello, the Moor of Venice*의 남성 주인공인 Othello와 그의 아내 Desdemona 사이의 대화코퍼스를 기반으로 VADER 감정 사전을 활용하여 각 대사에 해당하는 단어들의 값을 알아내고 기존의 감정 사전인 AFINN을 활용한 단어 점수들과도 비교해 본다. 4.4.1.3에서는 *Antony and Cleopatra*의 남성 주인공인 Antony 와 여성 주인공인 Cleopatra 사이의 대화코퍼스를 기반으로 VADER 감정 사전을 활용하여 각 대사에 해당하는 단어들의 값을 알아내고 기존의 감정 사전인 AFINN을 활용한 단어 점수들과도 비교해 본다.

본 소설에서 최대 +4, 최저 -4의 점수 폭을 갖는 VADER 감정 점수와 최대 +5, 최저 -5의 점수 폭을 갖는 AFINN의 감정 점수를 정규화하여 두 감정 점수를 비교한 것은 아래 <Table 28>과 같다.

<Table 28> Normalization for VADER and AFINN Valences

VADER Valences		AFINN Valences	
-4	-1.00	-5	-1.00
-3	-0.75	-4	-0.80
-2	-0.50	-3	-0.60
...		...	
0	0	0	0
...		...	
+2	+0.50	+3	+0.60
+3	+0.75	+4	+0.80
+4	+1.00	+5	+1.00

위 <Table 28>에서 보는 바와 같이 두 감정 점수 모두가 -1에서 +1의 수치로

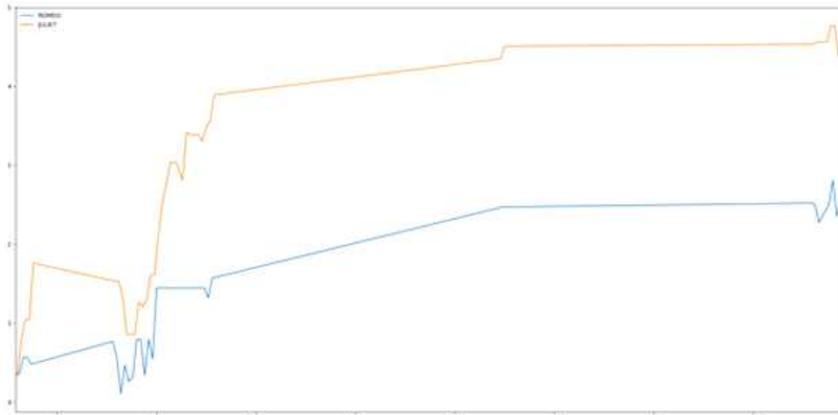
정규화된다. 따라서 VADER 감정 점수는 0.25 수치로 점수 폭을 정규화 시키고 AFINN 감정 점수는 0.20의 수치로 점수 폭을 정규화시켜 서로 다른 두 점수를 같은 점수 폭으로 비교할 수 있다.

#### 4.4.1.1 *Romeo and Juliet*

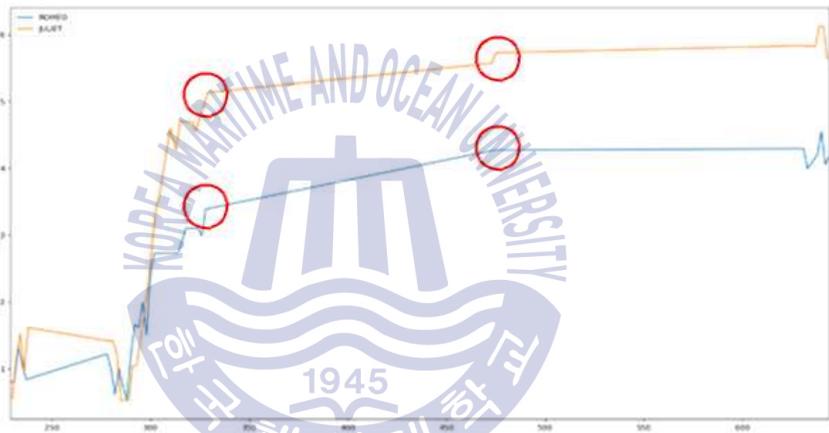
남자주인공인 Romeo와 여자주인공인 Juliet의 대화코퍼스를 이전 연구에서 활용된 AFINN 감정 사전으로 도출해낸 결과와 본 연구에서 활용한 VADER 감정 사전으로 도출해낸 결과를 비교한 것은 아래 <Fig. 15>와 같다.



AFINN



VADER



<Fig. 15> Comparing Cumulative Sentiment Valences using AFINN and VADER of Romeo-Juliet

위 <Fig. 15>에서 AFINN 사전과 VADER 사전을 기반으로 분석해낸 감정 분석이 비슷한 양상을 보이고 있다. 극의 초반부에서는 급격하게 상승하고 극의 후반부로 갈수록 일정하게 유지된다. 그러나 AFINN 사전을 기반으로 분석한 감성 분석은 극의 후반부에서 등장인물 Romeo와 Juliet의 그래프 폭이 극의 중반부와 비교했을 때 굉장히 좁은 것을 확인할 수 있다. 그리고 VADER 사전을 기반으로 분석한 감정 분석에서는 두 등장인물의 그래프 폭이 극의 중반부와 후반부에서 거의 동일한 폭의 크기를 보여주고 있다. 그래프에 가장 큰 차이를

보이는 부분의 시작점과 끝점의 Romeo-Juliet의 대사에 해당하는 단어들의 VADER 값과 AFINN값을 그래프에 직접 표시하였고 시작점과 끝점의 대사 누적 값을 살펴본다. 두 개의 그래프에서 가장 큰 차이를 보이는 부분의 시작점의 대사에 해당하는 단어의 VADER, AFINN 값을 나타낸 표는 아래 <Table 29 (a), (b)>와 같다.

<Table 29 (a)> Comparing Valences of VADER and AFINN of Romeo-Juliet

N	Act (Scene)	Speaker	Dialogue	Listener	Word	VADER	AFINN
329	2(2)	ROMEO	I wish I was your bird.	JULIET	wish	1.7	1.0
<b>Total</b>						1.7	1.0

<Table 29 (b)> Comparing Valences of VADER and AFINN of Juliet-Romeo

N	Act (Scene)	Speaker	Dialogue	Listener	Word	VADER	AFINN
330	2(2)	JULIET	My sweet, so do I. But I would kill you by petting you too much. Good night, good night. Parting is such sweet sorrow that I' ll say good night until tonight becomes tomorrow.	ROMEO	sweet(2)	2(4.0)	2(4.0)
					kill	-3.7	
					good(3)	1.9(5.7)	3(9.0)
					sorrow	-2.4	-2.0
<b>Total</b>						3.6	11.0

위 <Table 29 (a), (b)>에서 하나의 대사에서 여러 번 등장한 단어는 ()를 이용하여 반복된 횟수를 기입하였고 점수에서도 ()를 이용하여 그 반복된 횟수만큼 더한 점수를 기입하였다. 한 단어에 대해 두 사전에서 주어진 단어의 점수가 같은 단어가 있는가 하면 다른 점수가 부여되기도 하고 VADER사전에는 존재

하지만 AFINN사전에는 존재하지 않는 단어가 있었다. 예를 들어, ‘sweet’ 은 두 사전 모두 +2.0이라는 점수를 부여했고, ‘wish’ 는 VADER에서는 +1.7, AFINN에서는 +1.0으로, ‘good’ 은 VADER에서는 +1.9, AFINN에서는 +3.0으로 두 단어에서 부여되는 점수의 차이가 1 이상인 단어도 존재했다. ‘sorrow’ 는 VADER에서는 -2.4, AFINN에서는 -2.0의 점수를 부여 받았다. 또 ‘kill’ 과 같이 -3.7의 점수로 VADER사전에만 존재하는 단어도 찾아 볼 수 있었다.

두 개의 그래프에서 가장 큰 차이를 보이는 부분의 끝점의 대사에 해당하는 단어의 VADER, AFINN 값을 나타낸 표는 아래 <Table 30 (a), (b)>와 같다.

<Table 30 (a)> Comparing Valences of VADER and AFINN of Romeo-Juliet

N	Act (Scene)	Speaker	Dialogue	Listener	Word	VADER	AFINN
474	2(6)	ROMEO	Ah, Juliet if you’ re as happy as I am, and you’ re better with words, tell me about the happiness you imagine we’ ll have in our marriage.	JULIET	better	1.9	2.0
					happy	2.7	3.0
					happiness	2.6	3.0
<b>Total</b>						7.2	8.0

<Table 30 (b)> Comparing Valences of VADER and AFINN of Juliet-Romeo

N	Act (Scene)	Speaker	Dialogue	Listener	Word	VADER	AFINN
475	2(6)	JULIET	I’ ll give him equal thanks, so we’ re even.	ROMEO	thanks	1.9	2.0
<b>Total</b>						1.9	2.0

위 <Table 30 (a), (b)>에서는 한 단어에 대하여 두 사전에서 각각 다른 점수가 매겨진 사실을 발견 하였다. ‘better’ 는 VADER에서는 +1.9, AFINN에서는 +2.0으로, ‘happy’ 처럼 VADER에서는 +2.7, AFINN에서는 +3.0으로, ‘happiness’ 처럼 VADER에서는 +2.6, AFINN에서는 +3.0으로, ‘thanks’ 는 VADER에서는 +1.9, AFINN에서는 +2.0의 점수를 부여 받았다.

Romeo와 Juliet이 서로에게 주고받은 대사를 번호로 매기고 전체 대사 번호와 비교하여 각 대사의 VADER, AFINN값과 각 대사의 누적된 VADER, AFINN 값을 정리하고 얼마나 차이가 나는지 정리한 표는 아래 <Table 31>와 같다.

<Table 31> VADER and AFINN Valences and Cumulative VADER and AFINN Valences of Romeo-Juliet

N1	N2	Act (Scene)	Dialogue	VADER	AFINN 감정 점수	점수 차	누적 VADER 감정 점수	누적 AFINN 감정 점수	점수 차
230	1	1(5)	Your hand is like a holy place that my hand is unworthy to visit. If you’ re offended by the touch of my hand, my two lips are standing here like blushing pilgrims, ready to make things better with a kiss.	0.8037	0.3596	0.4441	0.8037	0.3592	0.4445
...									
328	53	2(2)	It’ s almost morning. I want to make you go, but I’ d only let you go as far as a spoiled child lets his pet bird go. He lets the bird hop a little from his hand and then yanks	0.1620	0.0833	0.0787	7.9130	4.8859	3.0271

			him back by a string.						
329	54	2(2)	I wish I was your bird.	0.4019	0.2500	0.1519	8.3149	5.1359	3.1790
330	55	2(2)	My sweet, so do I. But I would kill you by petting you too much. Good night, good night. Parting is such sweet sorrow that I' ll say good night until tonight becomes tomorrow.	0.2074	0.3247	-0.1173	8.5223	5.4606	3.0617
474	56	2(6)	I' ll give him equal thanks, so we' re even.	0.8807	0.4588	0.4219	8.9627	5.9194	3.0433
475	57	2(6)	Ah, Juliet if you' re as happy as I am, and you' re better with words, tell me about the happiness you imagine we' ll have in our marriage.	0.4404	0.9001	-0.4597	9.8434	6.8195	3.0239
...									
645	72	3(5)	And trust me, love, you look pale to me too. Sadness takes away our color. Goodbye, Goodbye!	0.1257	0.2395	-0.1138	9.8100	6.9664	2.8436

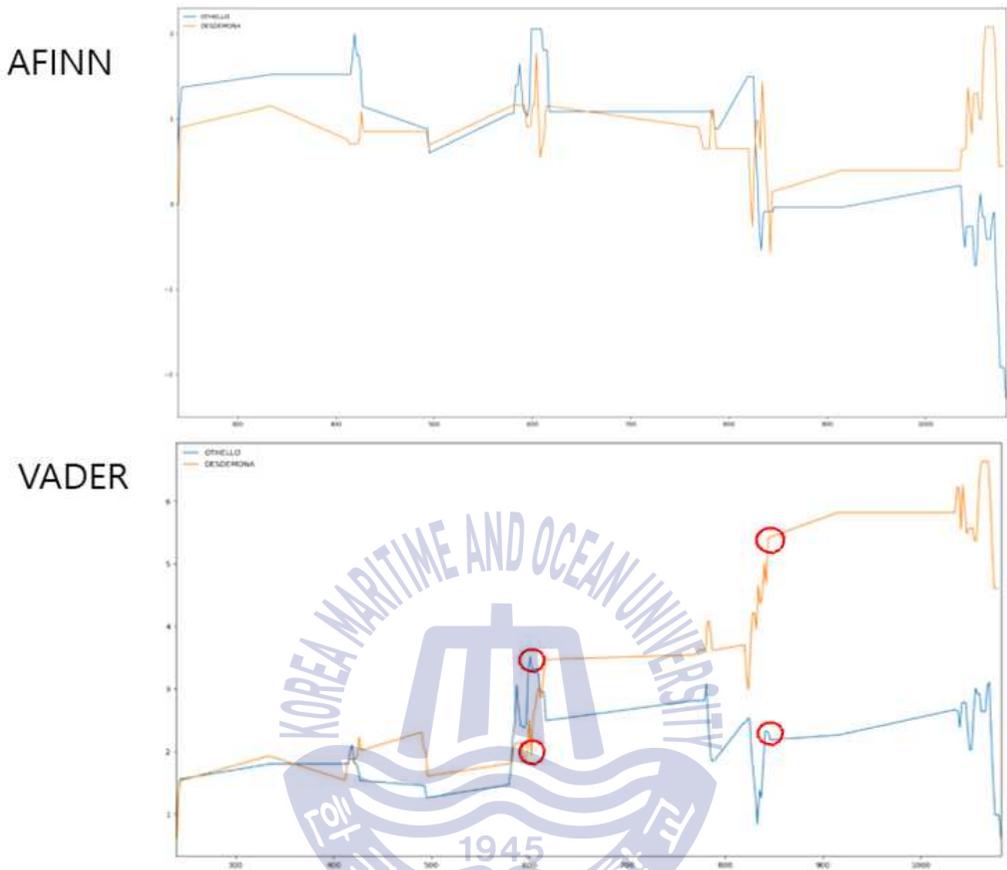
위 <Table 31>에서 번호 1은 극 *Romeo and Juliet*에서 대사 순서대로 번호를 매긴 대사 번호이고, 번호 2는 Romeo와 Juliet 두 사람 사이의 대화를 순서대로 번호를 매긴 대사 번호 이다. 앞서 언급한 두 개의 그래프에서 큰 차이를 보이는 부분의 시작점과 마지막 점의 Romeo-Juliet의 대화와 그 앞 뒤 대화도 연결하여 표에 나타냈다. Romeo와 Juliet, 서로의 대화에서 대사 번호가 1번 일 때 전체 극에서 볼 때는 230번이 되고 서로의 대화가 대사 번호 72번에서 끝나고 전체 극에서 볼 때는 645번이 된다. 또 대사의 VADER값과 AFINN값을 표기하고 누적값도 따로 표기하여 그 차이 값을 알아보았다. 일반적으로 VADER 누적

값이 AFINN 누적 값보다 높았으며 가장 큰 차이를 보이는 대사는 2막 2장에서 Romeo가 Juliet에게 한 대사로 3.1790의 차이를 보인다.

따라서 <Table 29 (a), (b)>, <Table 30 (a), (b)> 그리고 <Table 31>을 통해 기존의 연구 방법에 활용되었던 AFINN 감정 사전 보다 본 연구에서 활용된 VADER 감정 사전에서 가진 단어가 더 많았다. 예를 들어, 한 단어 대해 두 감정 사전 모두가 같은 점수를 부여하기도 하고 주어진 단어에 대해 서로 다른 점수를 부여하기도 했다. 또 ‘kill’ 과 같은 단어는 VADER 감정 사전에는 존재하지만 AFINN 감정 사전에는 존재하지 않는다. 따라서 VADER 감정 사전에는 더 많은 단어가 존재하고 그에 따라 더 세부적으로 평가 할 수 있다. 단어에 대한 점수를 정규화하여 VADER값과 AFINN값으로 나타내었을 때 그 차이가 크게 나타나며 그 점수들에 대한 누적 값으로 그래프 그려 결과를 관찰하였을 때 큰 폭으로 차이가 나는 것을 확인하였다.

#### 4.4.1.2 *Othello, the Moor of Venice*

극 *Othello, the Moor of Venice*에 등장하는 Othello와 Desdemona의 대화코퍼스를 이전 연구에서 활용된 AFINN 감정 사전과 본 연구에서 활용한 VADER 감정 사전으로 도출해낸 누적 감정 점수 결과를 비교한 그래프는 아래 <Fig. 16>와 같다.



<Fig. 16> Comparing Cumulative Sentiment Valences using AFINN and VADER of Othello-Desdemona

위 <Fig. 16>을 통해 극의 후반부 마지막 부분에서 남자주인공인 Othello와 여자주인공인 Desdemona의 누적 AFINN 감정값이 극명하게 차이가 나는 것을 확인할 수 있고 VADER 사전 기반으로 분석한 결과 극의 중반부를 기점으로 두 주인공의 누적 VADER 감정 점수가 반전되는 것을 관찰하였다. AFINN감정사전을 통한 감정 분석 그래프에서 Othello는 중반부를 기점으로 마이너스 VADER 감정 점수를 가진다. 그러나 Desdemona는 중반부를 기점으로 플러스 VADER 감정 점수를 가지며 큰 폭으로 증가하는 것을 확인할 수 있다.

아래의 표들은 그래프에 직접 표시한 가장 큰 차이를 보이는 부분의 시작점과 끝점의 Othello와 Desdemona의 대사에 해당하는 단어들의 VADER 값과 AFINN값을 정리한 것이다. 표를 통해 두 부분의 대사의 누적 값 변화를 살펴본다. 두 개의 그래프에서 가장 큰 차이를 보이는 부분의 시작점의 대사에 해당하는 단어의 VADER, AFINN 값을 나타낸 표는 아래 <Table 32 (a), (b)>과 같다.

<Table 32 (a)> Comparing Valences of VADER and AFINN of Othello-Desdemona

N	Act (Scene)	Speaker	Dialogue	Listener	Word	VADER	AFINN
604	3(4)	OTHELLO	Is it lost? Is it gone? Tell me, is it missing?	DESDEMONA	lost	-1.3	
					missing	-1.2	-2.0
<b>Total</b>						-2.5	-2.0

<Table 32 (b)> Comparing Valences of VADER and AFINN of Desdemona-Othello

N	Act (Scene)	Speaker	Dialogue	Listener	Word	VADER	AFINN
605	3(4)	DESDEMONA	God help me!	OTHELLO	god	1.1	
					help	1.7	2.0
<b>Total</b>						2.8	2

위 <Table 31 (a), (b)>에서는 한 단어에 대해 두 사전에서 주어진 점수가 다르게 나타난 유형과 하나의 사전에만 점수가 존재하는 유형을 살펴볼 수가 있다. 예를 들어 ‘missing’ 은 VADER에서는 -1.2, AFINN에서는 -2.0으로 주어졌고, ‘help’ 는 VADER에서는 +1.7, AFINN에서는 +2.0으로 주어졌다. 반면 ‘lost’

는 VADER사전에서만 -1.3의 점수가 주어졌고 AFINN에서는 주어진 점수가 없었다. 마찬가지로 ‘god’ 은 VADER사전에서만 +1.1이라는 점수가 주어졌다.

두 개의 그래프에서 가장 큰 차이를 보이는 부분의 마지막 점의 대사에 해당하는 단어의 VADER, AFINN 값을 나타낸 표는 아래 <Table 33 (a), (b)>와 같다.

<Table 33 (a)> Comparing Valences of VADER and AFINN of Othello – Desdemona

N	Act (Scene)	Speaker	Dialogue	Listener	Word	VADER	AFINN
840	4(2)	OTHELLO	So you’ re saying you’ re not a whore?	DESDEMONA	whore	-3.3	-4.0
<b>Total</b>						-3.3	-4.0

<Table 33 (b)> Comparing Valences of VADER and AFINN of Desdemona– Othello

N	Act (Scene)	Speaker	Dialogue	Listener	Word	VADER	AFINN
841	4(2)	DESDEMONA	No, I’ m as honest as I am Christian. If only letting my husband touch my body means I’ m not a whore, I’ m not a whore.	OTHELLO	no	-1.2	-1.0
					honest	2.3	
					whore(2)	-3.3	-4.0
<b>Total</b>						-5.5	-9.0

위의 <Table 33 (a), (b)>에서 살펴볼 수 있듯이 같은 단어에 대해 ‘whore’ 는 VADER에서는 -3.3, AFINN에서는 -4.0으로, ‘no’ 는 VADER에서는 -1.2,

AFINN에서는 -1.0으로 두 단어에서 부여되는 점수가 다른 유형을 발견하였다. 또 ‘honest’ 와 같이 VADER사전에서만 +2.3의 점수로 존재하는 유형도 발견하였다.

Othello 와 Desdemona 서로에게 주고받는 대사를 번호로 매기고 전체 대사 번호와 비교하여 각 대사의 VADER, AFINN값과 각 대사의 누적된 VADER, AFINN값을 정리하고 얼마나 차이가 나는지 정리한 것은 아래 <Table 34>와 같다.

<Table 34> VADER and AFINN Valences and Cumulative VADER and AFINN Valences of Othello-Desdemona

N 1	N 2	Act (Scene)	Dialogue	VADER valences	AFINN valences	점수 차	Cumulative VADER valences	Cumulative AFINN valences	점수 차
240	1	2(1)	My beautiful warrior!	0.6360	0.6124	0.0236	0.6360	0.6124	0.0236
...									
603	57	3(4)	Why are you yelling at me so angrily?	-0.5598	0.0000	-0.5598	5.4630	3.2117	2.2513
604	58	3(4)	Is it lost? Is it gone? Tell me, is it missing?	-0.2047	0.0000	-0.2047	5.2583	3.2117	2.0466
605	59	3(4)	God help me!	0.6239	0.6124	0.0115	5.8822	3.8241	2.0581
...									
839	107	4(2)	I swear to God you' re accusing me wrongly!	0.1260	-0.6124	0.7384	5.7470	0.2769	5.4701
840	108	4(2)	So you' re saying you' re not a whore?	0.5334	0.3333	0.2001	6.2804	0.2769	6.0035

841	109	4(2)	No, I' m as honest as I am Christian. If only letting my husband touch my body means I' m not a whore, I' m not a whore.	0.5283	-0.3250	0.8533	6.8087	-0.0481	6.8568
...									
1083	167	5(2)	What' s that noise? Aren' t you dead yet? I know I' m cruel, but I' m merciful too. I don' t want you to linger in pain.	-0.3619	-0.3687	0.0068	5.2296	-1.8407	7.0703

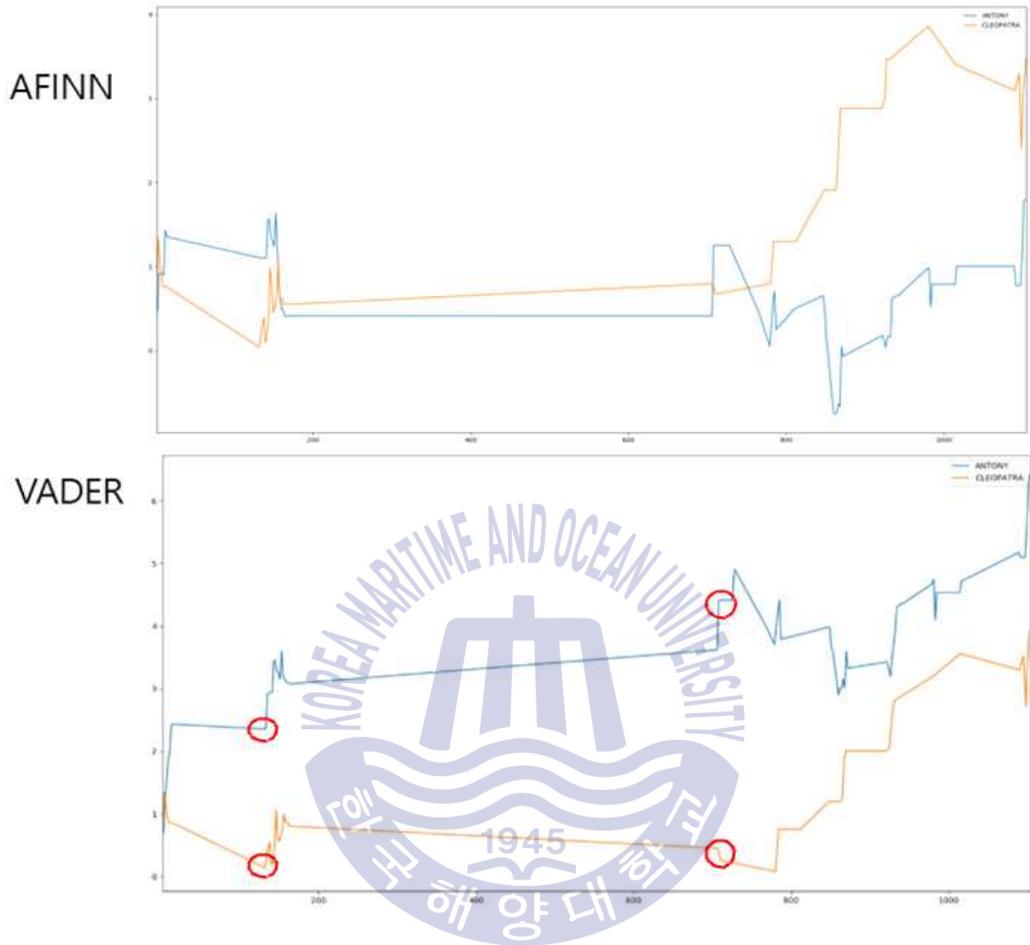
위의 <Table 34>에서 N 1은 극 *Othello, the Moor of Venice*에서 대사 순서대로 번호를 매긴 대사 번호이고, N 2는 Othello와 Desdemona 두 사람 사이의 대사를 순서대로 번호를 매긴 대사 번호이다. 따라서 Othello와 Desdemona, 서로의 대사는 대사 번호 1번으로 시작하지만 전체 극에서 볼 때는 240번이 된다. 서로의 대사 번호 167번에서 끝나고 전체 극에서 볼 때는 1,083번으로 끝이 난다. 두 개의 그래프에서 가장 큰 차이를 보이는 부분의 시작점과 마지막 점의 대사와 그 앞 뒤 대사도 포함시켰고 서로의 대사 처음 시작도 함께 표에 보여주었다. 그리고 대사의 VADER값과 AFINN값을 표기하고 누적 값도 따로 표기하여 그 차이를 알아보았다. 일반적으로 VADER 누적 값이 AFINN 누적 값보다 높았으며 가장 큰 차이를 보이는 대사는 5막 2장에서 Othello가 Desdemona에게 한 대사로 7.0703의 차이를 보인다. 이전 작품과 비교했을 때 VADER 누적 값과 AFINN 누적 값의 차이가 극의 흐름에 따라 점차 증가하는 것을 확인 할 수 있

다. 또한 N 2 기준의 대사번호 57, 58번에서는 AFINN 감정 점수가 0인 것을 확인하였다.

따라서 <Table 32 (a), (b)>, <Table 33 (a), (b)>, <Table 34>를 통해 기존의 연구 방법에 활용되었던 AFINN 감정 사전 보다 본 연구에서 활용된 VADER 감정 사전에서 가진 단어가 더 많다는 사실을 확인하였다. 예를 들어, ‘god’, ‘lost’ 과 같은 단어는 VADER 감정 사전에는 존재하지만 AFINN 감정 사전에는 존재하지 않고 주어진 단어에 대한 해당 점수들도 달랐다. 따라서 VADER 감정 사전에는 더 많은 단어가 존재하였고 그에 따라 더 세부적으로 평가 할 수 있었다. 단어에 대한 점수를 정규화하여 VADER 감정 점수와 AFINN 감정 점수로 나타내었을 때 그 차이가 크게 나타나며 각 점수들에 대한 누적 값으로 그래프 그려 결과를 관찰하였을 때 두 그래프의 폭이 점차 커지는 것을 확인 할 수 있다.

#### 4.4.1.3 *Antony and Cleopatra*

극 *Antony and Cleopatra*에 등장하는 Antony와 Cleopatra의 대화코퍼스를 이전 연구에서 활용된 AFINN 감정 사전으로 도출해낸 결과와 본 연구에서 활용한 VADER 감정 사전으로 도출해낸 결과를 비교한 그래프는 아래 <Fig. 17>과 같다.



<Fig. 17> Comparing Cumulative Sentiment Valences using AFINN and VADER of Antony–Cleopatra

위 <Fig. 17>은 AFINN 감정 사전으로 감정 분석한 그래프에서는 남자주인공 Antony와 여자주인공 Cleopatra의 누적 감정값의 차이가 작은 폭으로 나타나고 Cleopatra의 누적 감정 점수가 더 높은 것을 확인하였다. 그러나 VADER 감정 사전으로 감정 분석한 그래프에서는 큰 폭으로 차이가 났고 Othello의 누적 감정값이 더 크게 나타났다. 이후 극이 진행됨에 따라 후반부에서는 다양한 감정 변화를 보여주고 있다. Antony는 극의 후반부를 기점으로 AFINN값이 마이너스

로 급격하게 떨어지는 그래프를 보여주고 있다. 그와 반대로 Cleopatra는 극의 후반부를 기점으로 AFINN값이 급격하게 상승하는 것을 확인할 수 있다. 놀랍게도 VADER 사전 기반으로 감정 분석한 결과는 극이 진행됨에 AFINN 감정 분석과 상반된 모양의 그래프를 보여주고 있다. Antony는 극의 흐름에 따라 점차 상승하고 있다. 반면에 Cleopatra는 극의 초반부와 후반 시작 부분에서 Antony와는 다르게 하강하는 VADER값을 가지지만 극의 후반부를 기점으로 눈에 띄게 상승하는 VADER값을 가진다.

아래의 표들은 그래프에 직접 표시한 가장 큰 차이를 보이는 부분의 시작점과 마지막 점의 Antony와 Cleopatra의 대사에 해당하는 단어들의 VADER 감정 점수와 AFINN 감정 점수를 정리한 것이다. 표를 통해 두 부분의 대사의 누적 감정 점수 변화를 살펴본다.

두 개의 그래프에서 가장 큰 차이를 보이는 부분의 시작점 대사를 VADER 감정 사전과 AFINN 감정 사전의 단어 별 점수를 정리한 표는 아래 <Table 35 (a), (b)>와 같다.

<Table 35 (a)> Comparing Valences of VADER and AFINN of Cleopatra—Antony

N	Act (Scene)	Speaker	Dialogue	Listener	Word	VADER	AFINN
133	1(3)	CLEOPATRA	I am sick and sullen.	ANTONY	sick	-2.3	
					sullen	-1.7	-2.0
<b>Total</b>						-4.0	-2.0

<Table 35 (b)> Comparing Valences of VADER and AFINN of  
Antony–Cleopatra

N	Act (Scene)	Speaker	Dialogues	Listener	Word	VADER	AFINN
134	1(3)	ANTONY	I' m sorry to have to say this—	CLEOPATRA	sorry	-0.3	
<b>Total</b>						-0.3	

위 <Table 35 (a), (b)>를 통해 두 사전에서 한 단어에 대해 부여되는 점수가 다른 유형이 발견 되었고 하나의 사전에만 존재하는 유형도 발견되었다. 예를 들어, ‘sullen’ 은 VADER에서는 -1.7, AFINN에서는 -2.0으로 점수를 부여 받았고 ‘sick’ 는 VADER 사전에서만 -2.3의 점수를, ‘sorry’ 는 VADER 사전에서만 -0.3의 점수를 부여 받았다.

두 개의 그래프에서 가장 큰 차이를 보이는 부분의 끝점의 대사를 VADER 감정 사전과 AFINN 감정 사전의 단어 별 점수를 정리한 표는 아래 <Table 36 (a), (b)>와 같다.

<Table 36 (a)> Comparing Valences of VADER and AFINN of Antony–Cleopatra

N	Act (Scene)	Speaker	Dialogue	Listener	Word	VADER	AFINN
707	3(7)	ANTONY	Have you heard about this, sweet?	CLEOPATRA	sweet	2.0	2.0
<b>Total</b>						2.0	2.0

<Table 36 (b)> Comparing Valences of VADER and AFINN of Cleopatra–Antony

N	Act (Scene)	Speaker	Dialogue	Listener	Word	VADER	AFINN
708	3(7)	CLEOPATR A	No one admires speed more than the lazy.	ANTONY	no	-1.2	-1.0
					admires	1.5	3.0
					lazy	-1.5	-1.0
<b>Total</b>						-1.2	1.0

위 <Table 36 (a), (b)>를 통해 두 사전에서 한 단어에 대해 부여되는 점수가 같은 유형이 발견 되었고 한 단어에 대해 서로 다른 점수를 부여받은 유형도 발견되었다. 예를 들어, ‘sweet’ 은 두 사전 모두에서 +2.0의 점수를 부여 받았고, ‘no’ 는 VADER에서는 -1.2, AFINN에서는 -1.0의 점수를, ‘admires’ 는 VADER에서 +1.5, AFINN에서 +3.0의 점수를, ‘lazy’ 는 VADER에서 -1.5, AFINN에서 -1.0의 점수를 부여 받았다.

Othello 와 Desdemona 서로에게 주고받는 대화를 번호로 매기고 전체 대사 번호와 비교하여 각 대사의 VADER 값과 누적된 VADER 값을 정리한 표는 <Table 37>와 같다.

<Table 37> VADER and AFINN Valences and Cumulative VADER and AFINN Valences of Antony–Cleopatra

N1	N2	Act (Scene)	Dialogue	VADER valences	AFINN valences	점수 차	Cumulative VADER valences	Cumulative AFINN valences	점수 차
3	1	1(1)	If what you feel is really love, tell me how much.	0.6697	0.6124	0.0573	0.6697	0.6124	0.0573
...									
132	11	1(1)	Shame on you, stubborn Queen!	0.0087	-0.0781	0.0868	3.2859	2.1158	1.1701

			Everything you do is attractive—scolding, laughing, crying—every emotion seems admirable when you express it. I won’ t see any messengers but yours. Tonight we’ ll wander through the streets and observe the people. Come, my Queen. That’ s what you wanted to do last night.						
133	12	1(3)	I am sick and sullen.	-0.7184	-0.7184	0.0000	2.5678	1.3974	1.1704
134	13	1(3)	I’ m sorry to have to say this—	-0.0772	-0.2500	0.1728	2.4903	1.1474	1.3429
...									
706	41	1(3)	Let’ s go. Come with me. Our imminent separation so occupies our thoughts that even though you stay here, you come with me, and even though I leave here, I stay with you.	-0.0172	-0.0833	0.0661	3.8775	0.9608	2.9167
707	42	3(7)	Have you heard about this, sweet?	0.5423	0.0000	0.5423	4.4198	0.9608	3.4590
708	43	3(7)	No one admires speed more than the lazy.	-0.3535	0.2500	-0.6035	4.0663	1.2108	2.8555
...									
1105	109	4(15)	Most noble man, will you die? Don’ t you care about me? Shall I stay in this dreary world, which is no better than a pigsty without you?	0.0258	0.2226	-0.1968	10.3309	5.2852	5.0457

위 <Table 37>에서 N 1은 극 *Antony and Cleopatra*에서 대사 순서대로 번호를 매긴 대사 번호이고, N 2는 Antony와 Cleopatra 두 사람 사이의 대사를 순서대로 번호를 매긴 대사 번호이다. 따라서 Antony와 Cleopatra, 서로의 대사는 1번이지만 전체 극에서 볼 때 대사번호는 3번이 되고 서로의 대사에서 대사 번호 109번에서 끝나지만 전체 극에서 볼 때 1,105번이 된다.

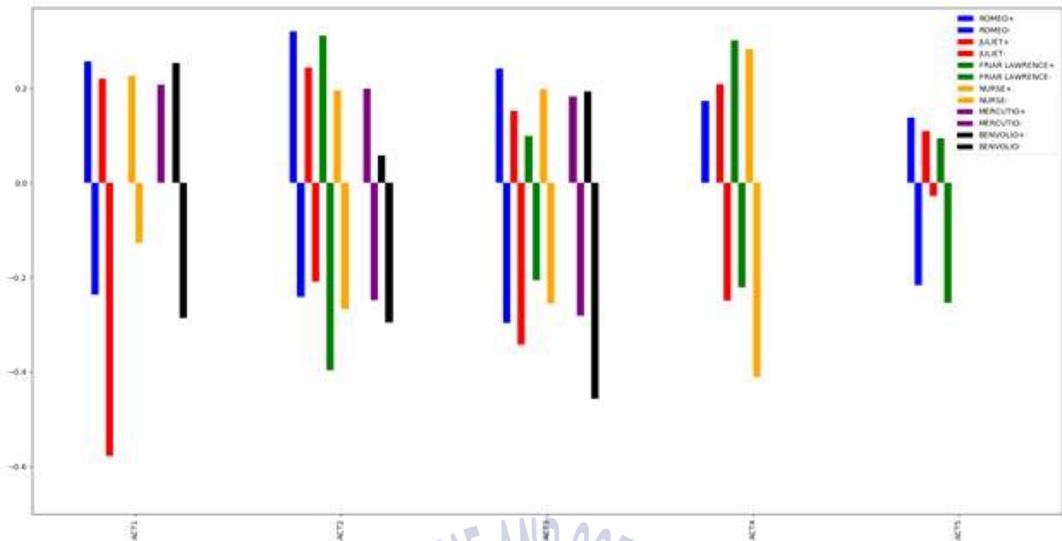
따라서 <Table 35 (a), (b)>, <Table 36 (a), (b)> 그리고 <Table 37>을 통해 기존의 연구 방법에 활용되었던 AFINN 감정 사전 보다 본 연구에서 활용된 VADER 감정 사전에서 가진 단어가 더 많았다. 예를 들어, ‘sick’, ‘sorry’와 같은 단어는 VADER 감정 사전에는 존재하지만 AFINN 감정 사전에는 존재하지 않는다. 따라서 VADER 감정 사전으로 많은 단어들을 더 세부적으로 평가할 수 있다는 사실을 확인하였다.

#### 4.4.2 막 별 주요 등장인물의 감정 분석

본 소절에서는 사랑 비극의 작품 별 주요 등장인물 간의 감정 분석을 한다. 즉, 여기서는 사랑 비극 세 편의 전체 대화코퍼스에서 STTR값을 가지는 주요 등장인물들의 감정 변화를 시간의 흐름을 나타내는 막 별로 구분하여 막대그래프의 시간시각화로 보여준다. 구체적으로 시간시각화의 막대그래프를 통해 주요 등장인물들의 극의 흐름에 따른 감정 변화의 특징을 기술하고 또한 각 작품의 막 내에서도 가장 긍정적인 인물과 부정적인 인물을 살펴본다.

##### 4.4.2.1 *Romeo and Juliet*

극 *Romeo and Juliet*에서 등장하는 각 주요 등장인물들의 감정 변화를 막의 흐름에 따라 어떻게 변화했는지를 나타낸 그래프는 아래 <Fig. 18>과 같다.



<Fig. 18> Comparison between All Major Characters in *Romeo and Juliet*

위 <Fig. 18>은 막을 기준으로 주요 등장인물들의 대화코퍼스를 VADER 감정 사전을 활용하여 감정 분석하여 감정 점수를 평균 내어 막대그래프로 보여주고 있다. 등장인물들의 감정 변화가 극의 흐름에 따라 다양하게 나타나고 감정 변화의 크기 또한 다양하게 나타난다.

등장인물 Romeo는 1막에서 5막까지 모두 등장한다. 그의 감정 그래프를 살펴보면 긍정적인 감정과 부정적인 감정이 항상 함께 공존한다. 1막과 2막에서는 긍정적인 감정 그래프의 값이 부정적인 감정의 값보다 크게 나타난다. 그러나 극의 후반부로 갈수록 부정적인 감정 그래프의 값이 더 커지고 있으므로 마지막 5막에서는 부정적인 감정이 더 큰 인물로 나타나고 있다.

두 번째 Juliet은 극의 초반인 1막에서는 눈에 띄게 부정적인 감정을 나타내고 있는 인물이다. 극의 흐름에 따라 Juliet은 부정적인 감정이 줄어들고, 3막에서는 다시 증가한다. 그러나 극의 후반부인 5막에서는 부정적인 감정 그래프보다 긍정적인 감정 그래프의 값이 더 크게 나타난다. 결과적으로 Juliet은 부정적인 감정을 가지고 있었지만 시간이 지남에 따라 긍정적인 감정을 가진 인물이라고 설명 할 수 있다.

세 번째 Friar Lawrence (Laurence) 는 두 주인공만큼이나 감정 변화가 심한 인물이다. 2막에서부터 등장하는데 3막에서는 부정적인 감정이 더 크게 나타난다. 4막에서는 긍정적인 감정 그래프의 값이 더 크지만 마지막 5막에서는 부정적인 감정 그래프가 훨씬 더 크게 나타난다. 따라서 Friar Lawrence는 결론적으로 부정적인 감정을 가진 인물이다.

네 번째로 Nurse는 극의 초반부 1막에서부터 등장하지만 결론 부분인 마지막 5막에는 등장 하지 않는다. 1막에서는 긍정적인 감정이 더 크게 나타나는 인물이었지만 극의 후반부로 갈수록 부정적인 감정 그래프 값이 더 크게 나타나는 인물이다.

작품 *Romeo and Juliet*에서는 두 가지 유형의 등장인물이 있음을 확인 할 수 있다. 먼저 극의 흐름에 따라 초반부에서는 부정적이었으나 후반부에서 긍정적인 감정을 더 크게 가지는 유형이 있고 다음으로 극의 흐름에 따라 긍정적인 인물에서 결론적으로 부정적인 인물이 되는 유형이 있다. 첫 번째 유형에는 Juliet이 해당된다. Juliet은 1막에서는 긍정적인 감정보다는 부정적인 감정 그래프가 더 크게 나타난다. 그러나 시간이 지남에 따라 극의 후반부에서는 긍정적인 감정 그래프가 우세하게 나타난다. 두 번째 유형에 해당하는 인물은 Romeo, Friar Lawrence, Nurse가 있다. 이들은 모두 극의 초반부에는 긍정적인 감정 그래프의 값이 더 크지만 마지막 후반부에서는 부정적인 감정 그래프의 값이 눈에 띄게 크게 나타난다.

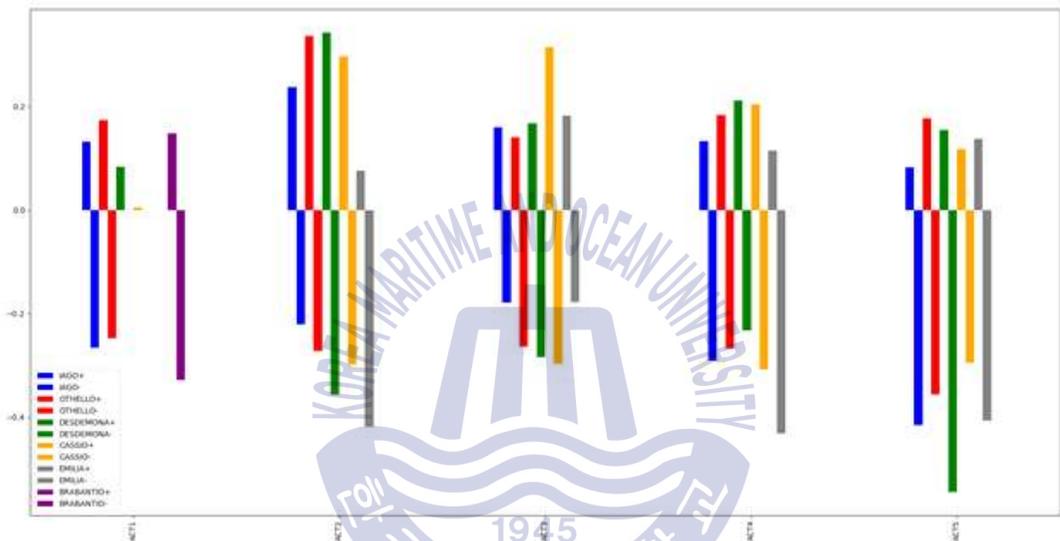
다음으로 막 별로 어떤 인물이 가장 긍정적인지 부정적인지를 살펴보았다. 1막에서 긍정적인 VADER값의 평균이 가장 높은 인물은 Romeo이다. 반면에 부정적인 VADER값의 평균이 가장 높은 인물은 Juliet이다. 그러나 막 내에서 긍정, 부정 VADER 평균값을 모두 반영한다면 Mercutio가 가장 긍정적인 인물이라고 할 수 있겠다. 2막에서는 긍정적인 VADER값의 평균이 가장 높은 인물은 Romeo이다. 반면에 부정적인 VADER값의 평균이 가장 높은 인물은 Friar Lawrence이다. 3막에서는 긍정적인 VADER값의 평균이 가장 높은 인물은 Romeo이다. 반면에 부정적인 VADER값의 평균이 가장 높은 인물은 Benvolio이다. 4막에서 긍정적인 VADER값의 평균이 가장 높은 인물은 Friar Lawrence이

다. 반면에 부정적인 VADER값의 평균이 가장 높은 인물은 Nurse이다. 그러나 막 내에서 긍정, 부정 VADER 평균값을 모두 반영한다면 Romeo가 가장 긍정적인 인물이라고 할 수 있겠다. 5막에서 긍정적인 VADER값의 평균이 가장 높은 인물은 Romeo이다. 반면에 부정적인 VADER값의 평균이 가장 높은 인물은 Friar Lawrence이다. 그러나 막 내에서 긍정, 부정 VADER 평균값을 모두 반영한다면 Juliet이 가장 긍정적인 인물이라고 할 수 있겠다.



#### 4.4.2.2 *Othello, the Moor of Venice*

작품 *Othello, the Moor of Venice*에서 등장하는 각 주요 등장인물들의 감정 변화를 막의 흐름에 따라 나타낸 그래프는 아래 <Fig. 19>와 같다.



<Fig. 19> Comparison between All Major Characters in *Othello, the Moor of Venice*

극 <Fig. 19>에서는 극의 흐름에 따라 등장인물들의 감정 변화가 다양하게 나타난다. 먼저 Othello는 1막에서 보다 2막에서 긍정적인 감정 그래프의 값이 크게 나타난다. 극의 초반부에서는 긍정적인 감정을 가진 인물이다. 그러나 극이 진행됨에 따라 긍정적인 감정 보다는 부정적인 감정 그래프 값이 더 크게 나타난다. Othello는 긍정적인 인물이었으나 시간이 지나면서 부정적으로 바뀐 인물이라고 설명 할 수 있다.

두 번째로 Desdemona는 1막에서는 긍정적인 감정 그래프만을 가진다. 2막에서도 여전히 긍정적인 감정이 더 크게 나타난다. 그러나 3막에서부터는 긍정적인 감정보다는 부정적인 감정이 훨씬 더 크게 나타나고 4막에서는 긍정적인 감

정과 부정적인 감정의 크기가 비슷하게 나타난다. 마지막 5막에서는 다른 어떤 등장인물보다도 부정적인 감정 그래프의 값이 크게 나타난다.

세 번째로 Iago는 극의 초반부인 1막과 2막에서는 긍정적인 감정을 보이다가 3막에서는 긍정적인 감정과 부정적인 감정의 크기가 비슷하게 나타나고 4막에서는 부정적인 감정값이 점차 커지기 시작한다. 마지막 극의 후반부인 5막에서는 부정적인 감정이 눈에 띄게 크게 나타난다는 것을 확인할 수 있다. 따라서 Iago는 점차적으로 부정적인 감정이 커지는 인물이라고 할 수 있다.

네 번째로 Cassio는 1막에서는 긍정적인 감정만을 가지지만 그 크기가 아주 미미하게 나타났다. 2막에서 긍정적인 감정이 점차 커지는데 극의 후반부인 4막과 5막으로 갈수록 긍정적인 감정 그래프는 줄어들지만 부정적인 감정 그래프는 평균적으로 비슷하게 나타난다. 따라서 긍정적인 감정이 줄어들어 결론적으로 부정적인 인물로 설명 할 수 있다.

다섯 번째 Emilia는 2막에서 처음 등장하는데 긍정적인 감정보다 부정적인 감정이 눈에 띄게 크게 나타난다. 3막에서는 긍정적인 감정이 더 크게 나타나지만 극이 후반부로 갈수록 긍정적인 감정보다는 부정적인 감정값이 굉장히 크게 나타난다. 따라서 Emilia는 3막을 제외한 나머지 2막, 4막, 5막에서 부정적인 값이 훨씬 더 크게 나타났으므로 결론적으로는 부정적인 인물이라고 할 수 있다.

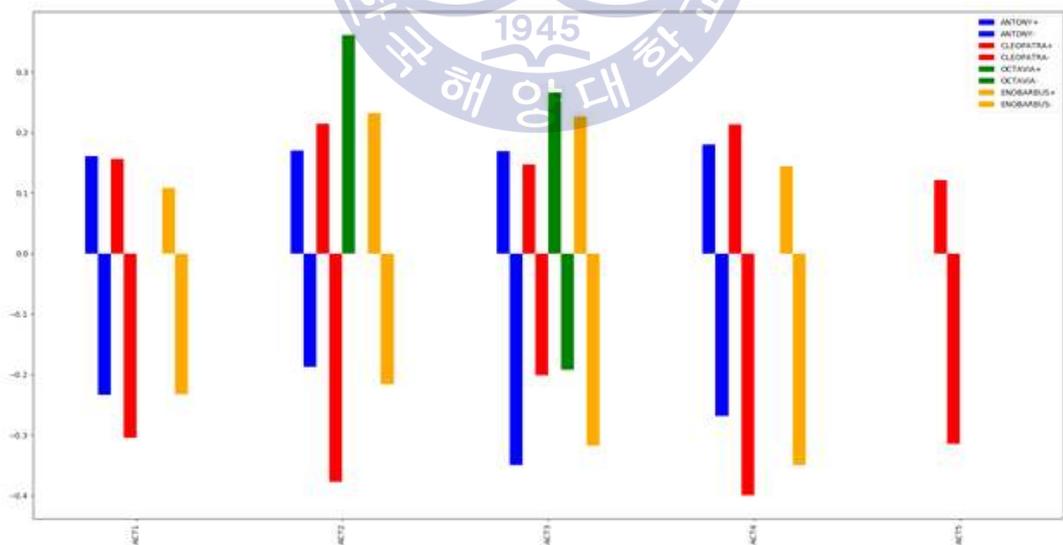
위 <Fig. 19>의 그래프를 통해 극 *Othello, the Moor of Venice*에서 Desdemona는 1막에서는 긍정적인 감정 그래프가 부정적인 감정 그래프보다 더 크게 나타난다. 시간이 지남에 따라 다른 등장인물들 보다 부정적인 감정 그래프가 가장 큰 것을 확인할 수 있으며 따라서 가장 부정적인 인물로 변화하였다. 그리고 극에 등장하는 등장인물들 모두가 극의 흐름에 따라 긍정적인 인물에서 결론적으로 부정적인 인물이 되는 유형에 속한다고 할 수 있다.

막 별로 어떤 인물이 가장 긍정적인지 또는 부정적인지를 살펴보고자 한다. 1막에서 긍정적인 VADER값의 평균이 가장 높은 인물은 Othello이다. 반면에 부정적인 VADER값의 평균이 가장 높은 인물은 Brabantio이다. 그러나 막 내에서 긍정, 부정 VADER 평균값을 모두 반영한다면 Desdemona가 가장 긍정적인

인물이라고 할 수 있겠다. 2막에서는 긍정적인 VADER값의 평균이 가장 높은 인물은 Desdemona이다. 반면에 부정적인 VADER값의 평균이 가장 높은 인물은 Emilia이다. 3막에서는 긍정적인 VADER값의 평균이 가장 높은 인물은 Cassio이고 동시에 부정적인 VADER값의 평균이 가장 높은 인물도 Cassio이다. Act 4에서 긍정적인 VADER값의 평균이 가장 높은 인물은 Desdemona이다. 반면에 부정적인 VADER값의 평균이 가장 높은 인물은 Emilia이다. 5막에서 긍정적인 VADER값의 평균이 가장 높은 인물은 Othello이다. 반면에 부정적인 VADER값의 평균이 가장 높은 인물은 Emilia이다.

#### 4.4.2.3 *Antony and Cleopatra*

극 *Antony and Cleopatra*에서 등장하는 각 주요 등장인물들의 감정 변화를 막의 흐름에 따라 나타낸 그래프는 아래 <Fig. 20>과 같다.



<Fig. 20> Comparison between All Major Character of *Antony and Cleopatra*

위 <Fig. 20>은 앞서 살펴본 그래프들과 마찬가지로 등장인물들의 감정 변화가 다양하고 그 크기도 다양하게 나타난다. Antony는 극이 진행됨에 따라 부정적인 감정 그래프 값이 꾸준히 증가하지만 상대적으로 긍정적인 감정 그래프 값은 일정하게 나타난다. 3막에서는 긍정적인 감정 그래프 보다 부정적인 감정 그래프가 더 크게 나타나고 4막에서도 부정적인 감정 그래프의 값이 더 크게 나타난다. 그러나 5막에서는 나타나지 않는다. 이는 Antony가 극의 후반부에 Cleopatra가 자결했다는 오보를 듣고 자결하였기 때문에 후반부에서 등장할 수가 없다.

두 번째 Cleopatra는 여성 등장인물로 1막과 2막에서는 부정적인 감정이 더 크게 나타난다. 3막에서는 부정적인 감정의 크기가 줄어들긴 했지만 여전히 부정적인 감정값이 더 크게 나타난다. 극이 진행되면서 계속 부정적인 감정값이 크게 나타난다. 이를 통해 Cleopatra는 극의 초반부터 마지막까지 부정적인 인물로 나타나고 있다.

세 번째 Enobarbus는 1막에서 부정적인 감정값이 크게 나타나는 부정적인 인물이다. 2막에서는 긍정적인 감정값이 더 크다. 하지만 시간이 흐름에 따라 부정적인 감정 그래프의 값이 증가하고 이에 반해 긍정적인 감정값은 떨어진다. 따라서 결론적으로 부정적인 인물이라는 것을 확인할 수 있다.

위 <Fig. 20>의 그래프를 통해 작품 *Antony and Cleopatra*에서는 모든 등장인물들이 긍정적인 감정보다는 부정적인 감정값이 크게 나타난다. 즉 모든 등장인물들이 부정적인 인물임을 확인 할 수 있으며 또한 Cleopatra가 극의 초반부에 비해 후반부에서 긍정적인 감정 그래프가 줄어들어 결론적으로 가장 부정적인 인물임을 확인 할 수 있다. 극 *Antony and Cleopatra*에 등장하는 등장인물들은 모두가 극의 흐름에 따라 긍정적인 인물에서 결론적으로 부정적인 인물이 되는 유형에 속한다고 할 수 있다.

막 별로 어떤 인물이 가장 긍정적인지 또는 부정적인지를 살펴보고자 한다. 1막에서 긍정적인 VADER값의 평균이 가장 높은 인물은 Antony이다. 반면에 부정적인 VADER값의 평균이 가장 높은 인물은 Cleopatra이다. 2막에서는 긍정

적인 VADER값의 평균이 가장 높은 인물은 Octavia이다. 반면에 부정적인 VADER값의 평균이 가장 높은 인물은 Cleopatra이다. 3막에서는 긍정적인 VADER값의 평균이 가장 높은 인물은 Octavia이고 부정적인 VADER값의 평균이 가장 높은 인물은 Antony이다. 4막과 5막에서 긍정적인 VADER값의 평균이 가장 높은 인물은 Cleopatra이고 동시에 부정적인 VADER값의 평균이 가장 높은 인물도 Cleopatra이다.

#### 4.5 요약

사랑 비극의 남녀주인공의 감정 추이변화를 통하여 AFINN 사전 그래프와 VADER 사전 그래프를 비교한 결과 VADER 감정 사전이 AFINN 감정 사전보다 우월하다는 것을 기술하였다. 즉, 감정 추이의 차이가 극명하게 드러나는 부분의 시작점의 대사와 끝부분의 대사를 비교 관찰하여 차이가 나는 이유와 감정 사전에 포함되어 있지 않는 단어에 대해서 기술하고 어떤 감정 사전이 감정 분석에 더 적합한 지를 확인을 위하여 대사의 나타나는 단어의 감정 점수의 차이 뿐만 아니라 감정 사전에 포함되어 있지 않는 단어도 보여주었다. 다시 말하면 두 사전이 한 단어에 대해 가지는 두 가지 유형 모두를 발견할 수 있었다. 첫째는 한 단어가 두 사전에서 각각 다른 점수를 부여받은 유형이다. 둘째는 한 단어가 하나의 사전에만 존재하는 유형이다. 즉 *Romeo and Juliet*에서는 두 가지 유형을 모두 발견할 수 있었다. 첫 번째 유형인 ‘wish’, ‘good’, ‘sorrow’, ‘better’, ‘happy’, ‘happiness’, ‘thanks’ 까지 모두 7개의 단어가 있고, 두 단어의 점수 차이가 1이상인 단어도 존재하였다. 또 두 번째 유형에 속하는 ‘kill’은 VADER 사전에만 -3.7의 점수로 존재하였다. 즉 *Othello, the Moor of Venice*에서는 첫 번째 유형에 속하는 ‘missing’, ‘help’, ‘whore’, ‘no’ 까지 총 4개의 단어가 있었다. 두 번째 유형에 속하는 ‘lost’, ‘god’, ‘honest’는 VADER 사전에서만 각각 -1.3, 1.1, 2.3의 점수를 부여받았다. *Antony and Cleopatra*에서는 두 가지 유형을 모두 발견할

수 있었다. 첫 번째 유형에는 ‘sullen’, ‘no’, ‘admires’, ‘lazy’ 와 같은 단어들이 있었고, 점수 차이가 1이상인 단어가 있었다. 두 번째 유형에는 ‘sick’, ‘sorry’ 가 VADER사전에서만 각각 -2.3, -0.3의 점수가 주어졌다. 두 번째 유형의 단어에서 점수가 없는 단어들은 0으로 누적으로 계산되므로 해당 단어들과 그 단어들을 포함한 대사의 VADER, AFINN점수 값을 큰 차이가 없었지만 누적 VADER값과 누적 AFINN값의 차이는 상대적으로 크게 나타났다. 따라서 이러한 누적 점수들의 차이가 <Fig. 15>, <Fig. 16>, <Fig. 17>에서 확인할 수 있는 큰 차이를 만들어 냈다. 이를 통해 VADER 감정 사전이 기존 연구에서 사용되었던 AFINN 감정 사전보다 더 많은 단어들을 가지고 있다는 것을 알아냈다. 가지고 있는 단어가 많을수록 더 세세하게 점수를 매길 수 있는 더 나은 평가 방법이었다는 사실을 확인하였다.

다음으로 사랑 비극 세 편의 전체 대화코퍼스에서 STTR값을 가지는 주요 등장인물들의 감정 변화를 막 별로 구분하여 막대그래프로 보여주었다. 막대그래프를 통해 시간의 흐름에 따른 각 작품의 막 간 등장인물들의 감정 변화가 긍정적인지 부정적인지를 확인할 수 있었고 막 내에서 가장 긍정적인 인물과 부정적인 인물을 발견할 수 있었다. 세 작품의 등장인물들의 감정 변화를 살펴보면 극의 흐름에 따라 초반부에서는 부정적이었으나 후반부에서 긍정적인 감정을 더 크게 가지는 유형이 있고 다음으로 극의 흐름에 따라 긍정적인 인물에서 결론적으로 부정적인 인물이 되는 유형이 있다. 작품 *Romeo and Juliet*에서는 Juliet은 극의 초반에는 부정적인 감정그래프가 더 크게 나타난다. 그러나 시간이 지남에 따라 극의 후반부에서는 긍정적인 감정그래프가 우세하게 나타났다. Romeo, Friar Lawrence, Nurse는 모두 극의 초반부에는 긍정적인 감정 그래프의 값이 더 크지만 마지막 후반부에서는 부정적인 감정그래프의 값이 눈에 띄게 크게 나타난다. 따라서 Juliet은 부정적인 인물이었지만 결론적으로 긍정적인 인물로 평가되고 나머지 Romeo, Friar Lawrence, Nurse는 긍정적인 인물이었지만 결론적으로는 부정적인 인물로 평가된다. *Othello, the Moor of Venice*에서는 Desdemona가 가장 부정적인 인물로 변화해간다는 것을 확인할 수 있으며 동시에 가장 부정적인 인물임을 알 수 있다. 따라서 극 *Othello, the Moor of*

*Venice*에 등장하는 등장인물들은 모두가 극의 흐름에 따라 긍정적인 인물에서 결론적으로 부정적인 인물이 되는 유형에 속한다고 할 수 있다. 극 *Antony and Cleopatra*에서는 모든 등장인물들이 긍정적인 감정보다는 부정적인 감정값이 크게 나타난다. 즉 모든 등장인물들이 부정적인 인물임을 확인 할 수 있으며 또한 Cleopatra가 결론적으로 가장 부정적인 인물임을 확인 할 수 있다. 극 *Antony and Cleopatra*에 등장하는 등장인물들은 모두가 극의 흐름에 따라 긍정적인 인물에서 결론적으로 부정적인 인물이 되는 유형에 속한다고 할 수 있다.



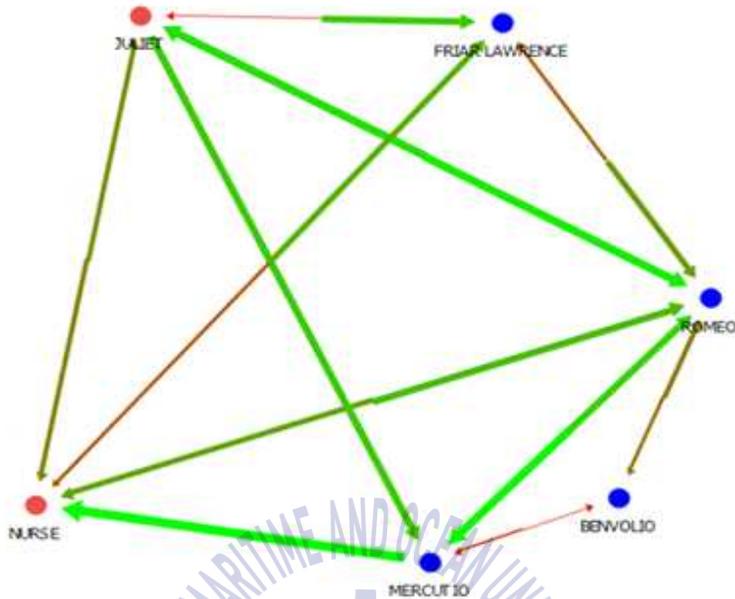
## 제 5 장 사랑 비극의 네트워크 분석

여기서는 세 편의 사랑 비극의 주요 등장인물들 간의 대화코퍼스에 나타난 대화의 방향성과 강도를 네트워크로 시각화하여 분석한다.

### 5.1 사랑 비극의 주요 등장인물의 네트워크 분석

#### 5.1.1 *Romeo and Juliet*

극 *Romeo and Juliet*의 대화코퍼스를 기반으로 주요 등장인물 간의 네트워크를 연결시켜 감정값의 강도를 시각화한 것은 아래 <Fig. 21>과 같다.



<Fig. 21> Network Analysis of Six Major Characters in *Romeo and Juliet*

위 <Fig. 21>에서 보이는 화살표의 붉은색이 짙을수록 더 부정적이고 초록색은 굵고 밝을수록 더 긍정적인 대화를 나누었다는 의미이다. 화살표의 굵기는 감정의 강도가 얼마나 부정적 또는 긍정적인지의 정도로 나타낸다. 또한 여성 등장인물은 붉은 색으로 남성 등장인물은 파란색으로 표시했다. 즉 *Romeo and Juliet*에 등장하는 주요 등장인물은 총 여섯 명이다. 그 중 여성 등장인물은 총 두 명으로 Juliet, Nurse가 있고 남성 등장인물은 Romeo, Friar Lawrence, Benvolio, Mercutio 총 네 명이 등장한다.

먼저 남자주인공인 Romeo와 여자주인공인 Juliet은 서로에게 긍정적인 네트워크를 가진다. 네트워크를 자세히 살펴보면 Romeo가 Juliet에게 가지는 네트워크보다 Juliet이 Romeo에게 가지는 네트워크가 더 긍정적인 것을 확인할 수 있다. 두 주인공 모두와 연결되어있는 주요 등장인물은 Friar Lawrence, Mercutio, Nurse가 있고 Benvolio는 Romeo와만 연결되어있다. Friar Lawrence는 Romeo에게는 약한 긍정적인 네트워크를 가지고 Romeo는 Friar Lawrence에게 약한 부

정적인 네트워크를 가진다. 반면 Juliet에게는 부정적인 네트워크를 가지고 Juliet은 Friar Lawrence에게 강한 긍정적인 네트워크를 가진다. 다시 말해 Friar Lawrence는 Romeo는 긍정적인 시각으로 바라보았고 Juliet은 부정적인 시각으로 바라보았다.

Mercutio는 Romeo에게 강한 긍정적인 네트워크를 가지고 Romeo도 Mercutio에게 강한 긍정적인 네트워크를 가진다. 한편 Juliet에게도 강한 긍정적인 네트워크를 가지고 Juliet도 Mercutio에게 강한 긍정적인 네트워크를 가진다. 즉, Mercutio는 Romeo와 Juliet 모두를 긍정적인 시각으로 바라보았다.

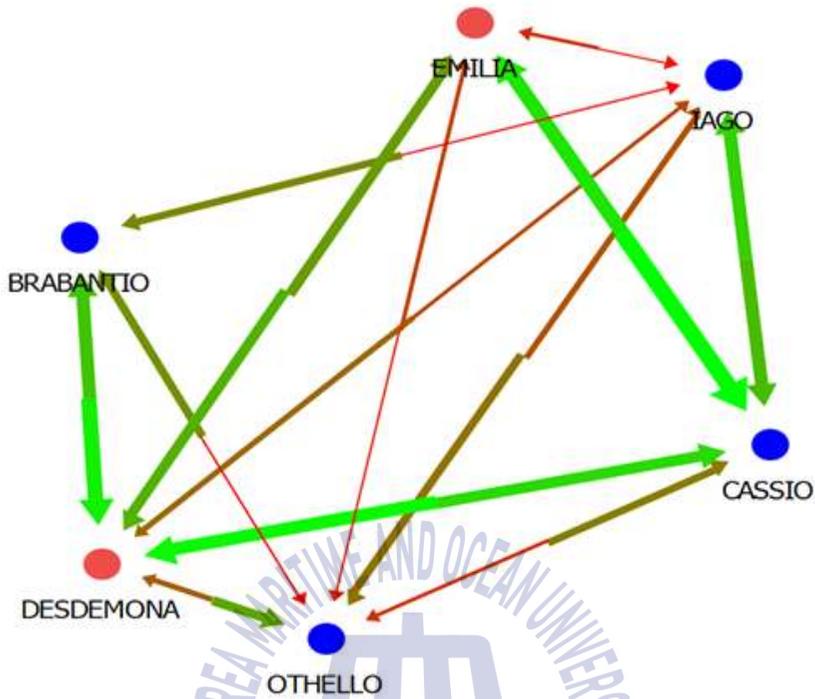
Nurse는 Romeo에게는 긍정적인 네트워크를 가지고 Romeo는 Nurse에게 약한 긍정적인 네트워크를 가진다. 반면 Juliet에게는 약한 긍정적인 네트워크를 가지고 Juliet도 Nurse에게 약한 긍정적인 네트워크를 가진다. 즉, Nurse는 Romeo와 Juliet 모두를 긍정적인 시각으로 바라보지만 Juliet보다 Romeo를 더 긍정적인 시각으로 바라보았다.

Benvolio는 유일하게 Juliet과의 네트워크를 가지지 않는 인물이다. 또한 Romeo에게 약한 긍정적인 네트워크를 가지고 Romeo도 Benvolio에게 약한 긍정적인 네트워크를 가진다. 즉, Benvolio는 Romeo를 긍정적인 시각으로 바라보았다.

지금까지 <Fig. 21>의 네트워크를 통해 주요 등장인물 여섯 명의 네트워크를 살펴보았다. 이를 통해 가장 긍정적인 네트워크로 연결된 등장인물은 Romeo와 Juliet, Romeo와 Mercutio, Juliet과 Mercutio이다. 또한 서로 가장 상반된 네트워크로 연결된 등장인물은 Juliet과 Friar Lawrence이다.

### 5.1.2 *Othello, the Moor of Venice*

극 *Othello, the Moor of Venice*에 등장하는 주요 인물들을 대화코퍼스를 기반으로 네트워크를 연결시켜 아래 <Fig. 22>로 보여준다.



<Fig. 22> Network Analysis of Six Major Characters in *Othello, the Moor of Venice*

위 <Fig. 22>에서 보이는 화살표의 붉은색이 얇을수록 더 부정적이고 초록색은 굵고 밝을수록 더 긍정적인 대화를 나누었다는 의미이다. 또 캐릭터의 성별을 한눈에 구별하기 위해 여성은 붉은색으로 남성은 파란색으로 표시했다.

극 *Othello, the Moor of Venice*에는 총 여섯 명의 주요 등장인물들이 있다. 그 중 여성은 총 두 명으로 Desdemona, Emilia이고 남성 등장인물은 총 네 명으로 Othello, Iago, Brabantio, Cassio이다.

먼저 남자주인공 Othello와 여자주인공 Desdemona의 네트워크를 살펴보면, 상반된 네트워크를 가진다는 것을 확인할 수 있다. 우선 Othello는 Desdemona에게 부정에 가까운 약한 긍정적인 네트워크를 가진다. 반면에 Desdemona는 Othello에게 긍정적인 네트워크를 가진다. 두 주인공과 나머지 주요 등장인물 Brabantio, Emilia, Iago, Cassio는 모두 연결되어있다.

Brabantio는 Othello에게는 강한 부정적인 네트워크를 가지고 Othello는 Brabantio에게 긍정적인 네트워크를 가진다. 반면 Desdemona에게는 강한 긍정적인 네트워크를 가지고 Desdemona도 Brabantio에게 강한 긍정적인 네트워크를 가진다. 또한 Brabantio가 Desdemona에게 가지는 네트워크가 Desdemona가 Brabantio에게 가지는 네트워크가 더 긍정적인 것을 확인할 수 있다. 즉, Brabantio는 Othello는 부정적인 시각으로 바라보았고 Desdemona는 굉장히 긍정적인 시각으로 바라보았다.

Emilia는 Othello에게 부정적인 네트워크를 가지고 Othello도 Emilia에게 강한 부정적인 네트워크를 가진다. 반면에 Desdemona에게는 긍정적인 네트워크를 가지고 Desdemona도 Emilia에게 긍정적인 네트워크를 가진다. 즉, Emilia는 Othello는 부정적인 시각으로 Desdemona는 긍정적인 시각으로 바라보았다.

Iago는 Othello를 부정적 가까운 긍정적인 네트워크를 가지고 Othello는 Iago에게 부정적인 네트워크를 가진다. 한편 Desdemona는 Iago에게 부정적인 네트워크를 가지지만 Iago는 Desdemona에게 부정적 가까운 긍정적인 네트워크를 가진다. 즉, Iago는 Othello와 Desdemona 모두에게 부정적 가까운 시각으로 바라보았다.

Cassio는 Othello에게 부정적인 네트워크를 가지지만 Othello는 Cassio에게 부정적 가까운 긍정적인 네트워크를 가진다. 반면 Desdemona에게는 강한 긍정적인 네트워크를 가지며 Desdemona도 Cassio에게 강한 긍정적인 네트워크를 가진다. 그러나 Cassio가 Desdemona에게 가지는 네트워크가 Desdemona가 Cassio에게 가지는 네트워크보다 더 긍정적이고 강한 것을 확인할 수 있다. 다시 말해, Cassio는 Othello를 부정적인 시각으로 바라보았고 Desdemona를 아주 강한 긍정적인 시각으로 바라보았다는 것을 확인할 수 있었다.

위 <Fig. 22>를 통해 극 *Othello, the Moor of Venice*의 네트워크를 살펴본 결과 가장 긍정적인 네트워크로 연결된 등장인물은 Desdemona와 Cassio, Desdemona와 Brabantio, Desdemona와 Emilia 순이다. 또한 가장 부정적인 네트워크로 연결된 등장인물은 Othello와 Emilia이다. 다음으로 서로 상반된 네트워

크로 연결된 등장인물은 Othello와 Cassio, Othello와 Brabantio, Othello와 Desdemona 순이다.

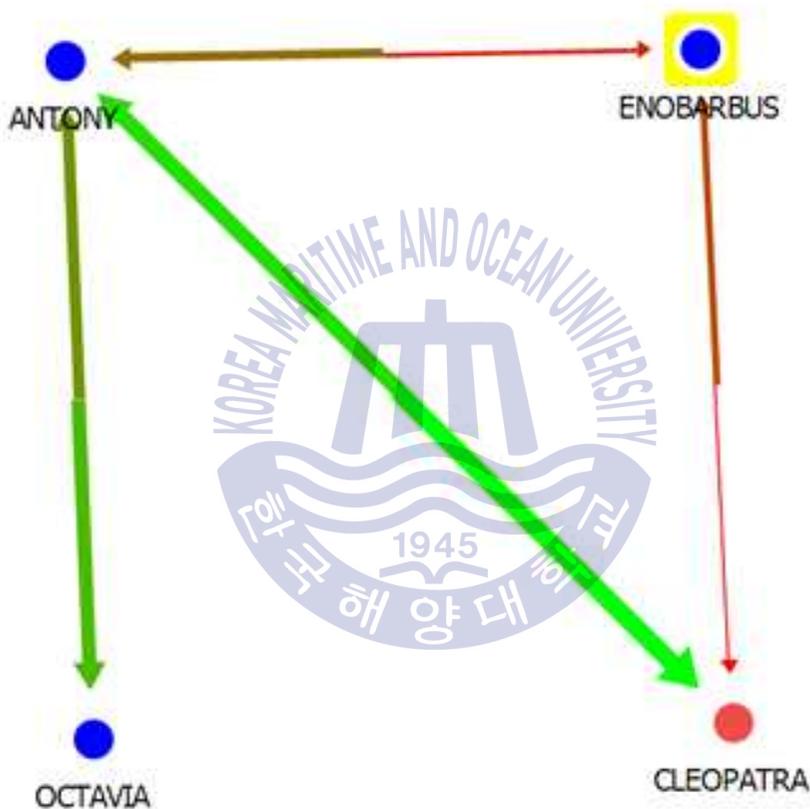
위 <Fig. 21>, <Fig. 22>과 마찬가지로 붉은 색 동그라미는 여성을 파란색 동그라미는 남성을 의미한다. 즉 *Antony and Cleopatra*에는 총 네 명의 중요등장 인물이 있다. 여성 등장인물은 한 명으로 Cleopatra 이고, Antony, Enobarbus, Octavia 세 명의 남성 등장인물들이 등장한다.

남자주인공인 Antony와 여자주인공인 Cleopatra의 네트워크를 살펴보면, Antony와 Cleopatra는 서로에게 강한 긍정적인 네트워크를 가진다. Antony가 Cleopatra에게 가지는 네트워크가 Cleopatra가 Antony에게 가지는 네트워크보다 더 긍정적인 것을 확인할 수 있다.

다음으로 두 주인공과 네트워크를 가지는 등장인물에는 Enobarbus가 있다. Enobarbus는 Antony에게 부정에 가까운 긍정적인 네트워크를 가지고 Antony는 Enobarbus에게 부정적인 네트워크를 가진다. 한편 Cleopatra에게 부정적인 네트워크를 가지고 Cleopatra도 Enobarbus에게 강한 부정적인 네트워크를 가진다. 다시 말해, Enobarbus는 Antony는 상대적으로 긍정적인 시각으로 바라보았고 Cleopatra는 상당히 부정적인 시각으로 바라보았다는 것을 발견하였다.

### 5.1.3 *Antony and Cleopatra*

극 *Antony and Cleopatra*의 주요 등장인물들의 대화코퍼스를 기반으로 네트워크를 연결시켜 아래 <Fig. 23>으로 보여준다.



<Fig. 23> Network Analysis of Four Major Characters in *Antony and Cleopatra*

위 <Fig. 23>을 통해 가장 긍정적인 네트워크로 연결을 가진 등장인물은 Antony와 Cleopatra이다. 반면 가장 부정적인 네트워크로 연결된 등장인물은 Cleopatra와 Enobarbus이다.

## 5.2 데이터 시각화의 비교

앞 4장에서 보여준 누적 선그래프와 막대그래프는 막을 기준으로 극의 흐름을 시간시각화로 전체 작품의 감정 분석과 남녀 주인공의 감정 분석 뿐 아니라 각 등장인물의 감정 추이를 쉽게 나타냈다. 반면에 본 5장에서 등장인물 간 감정관계를 기술하기 위하여 네트워크 분석을 하여 등장인물 간의 감정 연결과 그 정도를 시각화하였다.

아래의 <Table 38>은 시간시각화와 관계시각화의 비교에서 보는 바와 같이 시간시각화에서는 전체 작품 및 남녀 주인공의 감정 분석은 누적 선 그래프로 보여주었고, 각 등장인물의 감정 추이는 긍정과 부정으로 나누어 누적 막대그래프로 나타냈다. 네트워크의 관계시각화에서는 등장인물 간의 연결을 긍정 또는 부정의 정도로 표시하여 보여주었다.

<Table 38> Comparing between Time Visualization and Network Visualization

	전체 작품의 감정 추이	등장인물 간의 감정 차이	등장인물 간의 감정 추이	긍정/부정 감정의 연결 정도
Time Visualization	YES	YES	YES	NO
Network Visualization	?	?	?	YES

위 <Table 38>은 시간시각화를 통해 전체 작품의 감정 추이, 등장인물 간의 감정 차이, 등장인물 간의 감정 추이를 확인 할 수 있었다. 그러나 등장인물 간 감정이 긍정적으로 연결되어있는지 또는 부정적으로 연결되었는지, 그 정도에 대해서는 알 수 없었다. 그러므로 제 4장에서 보여주지 못한 주요 등장인물간의 긍정/부정 감정의 연결 정도는 제 5장에서 네트워크를 통한 관계시각화로 보여주었다. 본 논문에서는 시간시각화로 보여준 특징들을 네트워크 관계시각

화로 모두를 다 보여줄 수 있는지에 대해서는 본 논문의 범위를 벗어나므로 후속 연구가 필요하다.

### 5.3 요약

세 편의 비극 작품에 등장하는 주요 등장인물들을 중심으로 네트워크를 살펴 보았다. 각 비극 작품의 등장인물들 중 STTR값을 가지는 인물들만을 네트워크로 연결시켰다. STTR 값은 등장인물의 Token 수가 1,000 이상인 등장인물이 갖는다. 5.1.1에서는 극 *Romeo and Juliet*에서 네트워크를 많이 가지는 순서대로 나열하고 또 긍정적인 네트워크를 가지는 순서로 나열하여 기술하였다. Romeo, Juliet, Nurse 은 모두 5개의 네트워크를 가지면 가장 많은 네트워크를 가진 인물들이다. 그중에서 Romeo가 가진 5개의 네트워크는 모두 긍정적인 네트워크였다. 5.1.2에서는 극 *Othello, the Moor of Venice*에서 네트워크를 많이 가지는 순서와 긍정적인 네트워크를 가지는 순서로 나열하여 기술하였다. Othello, Desdemona, Iago가 5개의 네트워크를 가지고 Cassio가 4개의 네트워크를 가진다. 그러나 그 중에서 Cassio가 가진 4개의 네트워크는 모두 긍정적인 네트워크이고 나머지 Desdemona는 3개의 긍정적인 네트워크를, Othello는 2개, Iago는 1개의 긍정적인 네트워크를 가진다. 마지막으로 5.1.3에서는 극 *Antony and Cleopatra*에서 Antony는 3개의 네트워크를 Cleopatra는 2개, Enobarbus도 2개의 네트워크를 가진다. 그 중에서 Antony는 3개의 긍정적인 네트워크를, Cleopatra는 2개의 긍정적인 네트워크를 가지지만 Enobarbus는 모두 부정적인 네트워크를 가진다는 것을 알 수 있었다.

세 편의 비극 작품의 네트워크 분석을 통해서 등장인물이 긍정적으로 평가되는 인물인지 부정적으로 평가되는 인물인지 또는 작품 내에서 쉽게 보이지 않았던 등장인물들 간의 관계와 그 관계가 긍정적인지 또는 부정적인지를 한 눈에 확인 할 수 있었다.

## 제 6 장 결 론

### 6.1 요약

본 연구에서는 언어적 규칙기반 알고리즘이면서, 단순한 사전 기반의 기계적인 감정분석보다 인간 중심적인 감정분석이 가능한 Hutto and Gilbert (2014)가 제시한 VADER 감정분석 알고리즘을 사용하였다. 기존의 연구들과 달리 셰익스피어 비극 작품의 대화코퍼스를 막을 기준으로 작품 및 주요 등장인물의 감정의 변화와 추이변화 양상을 극의 흐름에 따라 관찰하여 분석하고 그에 따라 등장인물 간 네트워크를 시각화하여 살펴보았다. 또한 이전의 감정 분석으로는 알 수 없었던 비극과 희극의 차이점을 패턴으로 설명하였고 네트워크를 시각화하여 각 작품별 주요 등장인물끼리의 얼마나 강하게 긍정적인 또는 부정적인 감정으로 연결되어있는지를 발견 할 수 있었다.

제 4장에서 세 편의 비극 작품 각각의 AFINN 점수 그래프와 VADER 점수 그래프를 비교하여 AFINN 감정 사전과 VADER 감정 사전의 차이를 살펴 볼 수 있었다. 또한 차이가 극명하게 드러나는 부분의 시작점 대사와 끝부분 대사를 살펴보고 그 대사의 단어에 해당하는 점수를 살펴보고, 각 대사 내의 단어들에 해당하는 감정 사전 점수들을 비교하여 차이점을 발견하였다. 극 *Romeo and Juliet*의 남성 주인공인 Romeo 와 여성 주인공인 Juliet 사이의 대화코퍼스를 기준으로 VADER 감정 사전을 활용하여 각 대사에 해당하는 단어들의 값을 알아내고 기존의 감정 사전인 AFINN을 활용한 단어 점수들과도 비교해 보았다. 극 *Othello, the Moor of Venice*의 남성 주인공인 Othello와 그의 아내 Desdemona 사이의 대화코퍼스를 기본으로 VADER 감정 사전을 활용하여 각 대사에 해당하는 단어들의 값을 알아내고 기존의 감정 사전인 AFINN을 활용한 단어 점수들과도 비교해 보았다. 극 *Antony and Cleopatra*의 남성 주인공인

Antony 와 여성 주인공인 Cleopatra사이의 대화코퍼스를 기본으로 VADER 감정 사전을 활용하여 각 대사에 해당하는 단어들의 값을 알아내고 기존의 감정 사전인 AFINN을 활용한 단어 점수들과도 비교해 보았다. 이를 통해 VADER 감정 사전이 기존 연구에서 사용되었던 AFINN 감정 사전보다 더 많은 단어들에 대한 점수를 가지고 있다는 것을 알아냈다. 가지고 있는 단어가 많을수록 더 세세하게 점수를 매길 수 있는 더 나은 평가 방법이었다는 사실을 확인 하였다. 마지막으로 세 편의 비극작품에 등장하는 주요 등장인물의 막 별 감정 변화를 살펴보고 막에서 어떤 인물이 가장 긍정적인 VADER 평균값을 가지는지, 어떤 인물이 가장 부정적인 VADER 평균값을 가지는 살펴보았다. 그리고 막 내에서도 어떤 등장인물이 가장 긍정적인지 또는 부정적인지를 막 내 주요 등장 인물들의 VADER 점수의 평균값으로 살펴보았다.

제 5장을 통해 세 편의 비극 작품에 등장하는 주요 등장인물들을 중심으로 네트워크를 살펴보았다. 각 비극 작품의 등장인물들 중 STTR값을 가지는 인물들만을 네트워크로 연결시켰다. 세 편의 비극 작품의 네트워크 분석을 통해서 등장인물이 긍정적으로 평가되는 인물인지 부정적으로 평가되는 인물인지 또는 작품 내에서 쉽게 보이지 않았던 등장인물들 간의 관계와 그 관계가 긍정적인지 또는 부정적인지를 한 눈에 확인 할 수 있었다. 5장에서 언급한 바와 같이 시간시각화를 통해 전체 작품의 감정 추이, 등장인물 간의 감정 차이, 등장인물 간의 감정 추이를 확인 할 수 있었다. 그러나 등장인물 간 감정이 긍정적으로 연결되어있는지 또는 부정적으로 연결되었는지, 그 정도에 대해서는 알 수 없었다. 그러므로 제 4장에서 보여주지 못한 주요 등장인물간의 긍정/부정 감정의 연결 정도는 제 5장에서 네트워크를 통한 관계시각화로 보여주었다. 본 논문에서는 시간시각화로 보여준 특징들을 네트워크 관계시각화로 모두를 다 보여줄 수 있는지에 대해서는 더 조사할 필요가 있으며 본 논문의 범위를 벗어나므로 후속 연구가 필요하다.

## 6.2 향후 과제

본 연구를 통해 감정 분석과 같은 양적인 통계방법으로 비극과 희극의 차이점을 규명하여 비극과 희극의 유형을 범주화하고 이러한 갈등과 대립의 요소가 어떻게 비극적인 삶을 살게 되는지 아니면 행복한 결말을 얻게 되는지에 대한 알고리즘(세익스피어가 미리 설정해 놓았다고 볼 수 있는 공식)의 개발 가능성을 기대할 수 있다. 또한 본 연구의 결과는 빅데이터 분석, 4차 산업혁명으로 이어지는 디지털 기술의 발전에 영문학의 다양한 텍스트를 활용한 새로운 연구에 활용될 수 있을 것이며 또한 영문학의 다양한 텍스트를 활용한 새로운 연구에 활용될 수 있을 것으로 기대한다. 그리고 다른 언어로 쓰인 극이나 소설과 같은 다양한 장르에서의 활용할 수 있을 것이다.



## 참고 문헌

- 이경은, 장세은, 김재훈, 박호민, 2018. 감정분석과 네트워크 분석을 통한 셰익스피어 사랑주제의 비극작품과 희극작품의 비교 연구. *2018 KACL & ISLI Joint Conference Using "Big Data" Web-Based Corpora for Research and Teaching held in Kyung Hee University.*
- 장세은, 이수상, 송원문, 정해룡, 이성민, 김재훈, 2017. 텍스트 마이닝을 통한 셰익스피어 학술논문 영어초록 코퍼스의 토픽모델링 분석. *언어과학*, 24(4), pp.55-82.
- 장세은, 김재훈, 박호민, 2018. 셰익스피어 비극 작품과 주요 등장-인물의 대화 코퍼스를 활용한 감정분석과 네트워크 분석. *2018 KACL & ISLI Joint Conference Using "Big Data" Web-Based Corpora for Research and Teaching held in Kyung Hee University.*
- 하명정, 2013. 코퍼스에 기반한 문학텍스트 분석. *한국콘텐츠학회논문지*, 13(9), pp.440-447.
- Archer, D., Culpeper, J. and Rayson, P., 2005. Love-a familiar or a devil? An exploration of key domains in Shakespeare's Comedies and Tragedies. *AHRC ICT Methods Network Expert Seminar on Linguistics.* pp. 137-157.
- Busse, U., 2002. *Linguistic variation in the Shakespeare corpus: Morpho-syntactic variability of second person pronouns* (Vol. 106). Amsterdam-Philadelphia: John Benjamins.
- Culpeper, J., 2009. Keyness: Words, parts-of-speech and semantic categories

- in the character-talk of Shakespeare's Romeo and Juliet. *International Journal of Corpus Linguistics*, 14(1), pp.29-59
- Gilbert, C. and Hutto, E., 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. *In Eighth International Conference on Weblogs and Social Media (ICWSM-14)*. [Online] (Updated 20 April 2016) Available at <http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf>. [Accessed 10 April 2018]
- Grandjean, M., 2015. Network visualization: mapping Shakespeare's tragedies [Online] (Updated 23 Dec 2015) Available at: <http://www.martingrandjean.ch/network-visualization-shakespeare/> [Accessed 10 April 2018]
- Hota, S.R., Argamon, S. and Chung, R., 2006. Gender in Shakespeare: Automatic stylistics gender character classification using syntactic, lexical and lemma features. *Digital Humanities and Computer Science (DHCS)*, pp.100-106.
- Hussain, A., Sattar, S. and Afzal, M.T., 2015. Literature Review on Feature Identification in Sentiment Analysis. *International Journal of Computer Applications*, 132(3), pp.22-27.
- Islam, M.R. and Zibran, M.F., 2017, November. A comparison of dictionary building methods for sentiment analysis in software engineering text. *In Empirical Software Engineering and Measurement (ESEM), 2017 ACM/IEEE International Symposium on IEEE*. pp.478-479.
- Kim, S.M. and Hovy, E., 2004, August. Determining the sentiment of opinions. *In Proceedings of the 20th international conference on Computational Linguistics (p. 1367)*. Association for Computational Linguistics.
- Liu, S. and Jhang, S.E., 2018. Keywords and Key Clusters in Shkaespeare's 3

Love Tragedies and 3 Love Comedies, *2018 KACL & ISLI Joint Conference Using “Big Data” Web-Based Corpora for Research and Teaching held in Kyung Hee University.*

Mahlberg, M., 2007. A corpus stylistic perspective on Dickens’ Great Expectations. *Contemporary Stylistics*, pp.19-31.

Modha, J.S., Pandi, G.S. and Modha, S.J., 2013. Automatic sentiment analysis for unstructured data. *International journal of advanced research in computer science and software engineering*, 3(12), pp.91-97.

Mohammad, S., 2011, June. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. *In Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*. Association for Computational Linguistics. pp. 105-114.

Moretti, F., 2011. Network theory, plot analysis. Pamphlets of the Stanford Literary Lab. [Online] (Updated 01 May 2011) Available at: <http://litlab.stanford.edu/LiteraryLabPamphlet2A.Text.pdf> [Accessed 10 April 2018]

Nalisnick, E.T. and Baird, H.S., 2013, August. Extracting sentiment networks from Shakespeare’s plays. *In Document Analysis and Recognition (ICDAR), 2013 12th International Conference on IEEE*. pp. 758-762.

Nalisnick, E.T. and Baird, H.S., 2013. Character-to-character sentiment analysis in Shakespeare’s plays. *In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (Vol. 2, pp. 479-483).*

Nalisnick, E.T., 2013. *Automatic Methods for Tracking Sentiment Dynamics in Plays*. Degree of Master. Pennsylvania: Lehigh University.

- Nielsen, F.Å., 2011. A new ANEW: Evaluation of a word list for sentiment analysis in microblogs. *Proceedings of the ESWC 2011 Workshop on 'Making Sense of Microposts': Big things come in small packages*, pp.93-98.
- Scott, M. and Tribble, C., 2006. *Textual Patterns: Keyword and Corpus Analysis in Language Education*. Amsterdam: Benjamins.
- Scott, M. 2009. Key Cluster and Concgram Patterns in Shakespeare, a manuscript presented at CL 2009, Liverpool, [Online] (Updated 22 July 2009) Available at: [www.lexically.net/downloads/Lit](http://www.lexically.net/downloads/Lit). [Accessed 10 April 2018]
- Scott, M., 2015. *WordSmith Tools Manual*. Version 6.0. Liverpool: Lexical Analysis Software Ltd. [Online] (Updated Dec 2015) Available at: [www.lexically.net/downloads/version6/wordsmith6.pdf](http://www.lexically.net/downloads/version6/wordsmith6.pdf) [Accessed 10 April 2018]
- Stiller, J. and Hudson, M., 2005. Weak links and scene cliques within the small world of Shakespeare. *Journal of Cultural and Evolutionary Psychology*, 3(1), pp.57-73.
- Stiller, J., Nettle, D. and Dunbar, R.I., 2003. The small world of Shakespeare's plays. *Human Nature*, 14(4), pp.397-408.
- Tripathy, A. and Rath, S.K., 2017. Classification of sentiment of reviews using supervised machine learning techniques. *International Journal of Rough Sets and Data Analysis (IJRSDA)*, 4(1), pp.56-74.
- West, R., Paskov, H.S., Leskovec, J. and Potts, C., 2014. Exploiting social network structure for person-to-person sentiment analysis. *Transactions of the Association for Computational Linguistics (TACL)*, 2. pp.297-310.