

# 사건의 흐름 분석을 위한 동화의 재구성

안승주<sup>o</sup>, 박종철  
KAIST 전산학과  
{sjan<sup>o</sup>, park}@nlp.kaist.ac.kr

## Reproducing Fairy Tales for Plot Identification

Seungjoo An<sup>o</sup>, Jong C. Park  
Computer Science Department, KAIST

### 요 약

텍스트의 스토리를 자동으로 이해하기 위해 텍스트에서 기술된 사건(event)을 파악하고 이들을 조합하여 스토리가 어떻게 구성되어 있는지를 파악하는 연구들이 진행되어 왔다. 하지만 이는 스토리의 깊은 의미론적 이해를 요구하는 것 이외에도 텍스트마다 상황과 일어나는 사건들이 다양하기 때문에 언어 자원이 부족한 환경에서의 처리에는 한계가 있다. 이러한 문제는 사건들을 추상화 하여 단순하게 표현할 수 있다면 스토리 이해의 자연스러움을 저해하지 않고 해결 할 수 있다. 본 논문에서는 사건들의 추상화 과정을 위한 기초 연구로서 텍스트 속 등장인물이 행하거나 당하는 사건들을 추출하여 PMI기법을 통해 사건의 흐름을 파악하고 언어학적 단서를 참조하여 스토리 이해 과정에 누락될 수 있는 사건들을 추가하여 보완하였다. 이러한 접근을 통해 등장인물이 행할 수 있는 사건들을 재구성하여 단순화하는 방법을 제시한다.

주제어: Event Extraction, Event Abstraction, Story Understanding

### 1. 서론

텍스트의 스토리를 이해하려는 시도는 자연언어 처리 분야에서 오래전부터 연구되어 왔다. 이러한 텍스트의 스토리 이해는 스토리의 요약(summarization)이나 질의 응답(question answering)으로부터 텍스트 애니메이션이나 자동 스토리 생성 시스템(story generation)과 같은 응용분야에 이르기까지 다양한 분야에 적용된다. 특히 텍스트의 내용을 이해함에 있어서 텍스트 내에서 벌어지는 사건(event)들을 정확하게 파악하는 것이 중요한데, 이러한 실제적인 사건들을 통하여 어떠한 흐름으로 내용이 전개되는지를 잘 이해할 수 있기 때문이다.

그렇지만 자연어처리 기법을 사용하여 스토리를 완전하게 이해하고자 하는 것은 어려운 문제이다. 텍스트마다 동일한 내용이더라도 텍스트마다 기술 방법이나 분량에 따라서 다양한 상황과 사건이 존재할 수 있기에, 텍스트 애니메이션과 같이 등장인물을 위한 리소스(resource)가 제한된 환경에서는 텍스트 내용을 항상 원문의 의도에 맞게 표현하는 것은 매우 어렵다.

텍스트 내에서 스토리는 처음부터 끝까지 연결된 하나의 흐름으로 간주될 수 있지만, 여러 개의 플롯(plot)들의 집합으로 구성되어 있다고 볼 수 있으며 이러한 각각의 플롯이 체인(chain)의 형태를 이루고 있다고도 볼 수 있다. 또한 각각의 플롯은 여러 개의 사건들로 구성될 수 있다. 실제로는 많은 행동과 상황이 묘사된 글일지라도 그림 1과 같이 추상화(abstraction)의 과정을 거치면

보다 단순한 단계(level)들로 계층화할 수 있고, 이렇게 하면 제한된 리소스 환경에서도 어느 정도 의미 있게 텍스트를 표현 할 수 있게 한다. 또한 이러한 스토리의 추상화 단계를 통하여 실제로는 전혀 다른 내용의 텍스트 일지라도 유사한 구조나 내용 전개의 흐름을 가질 수 있기에 텍스트간의 유사성 비교에도 도움을 줄 수 있다.

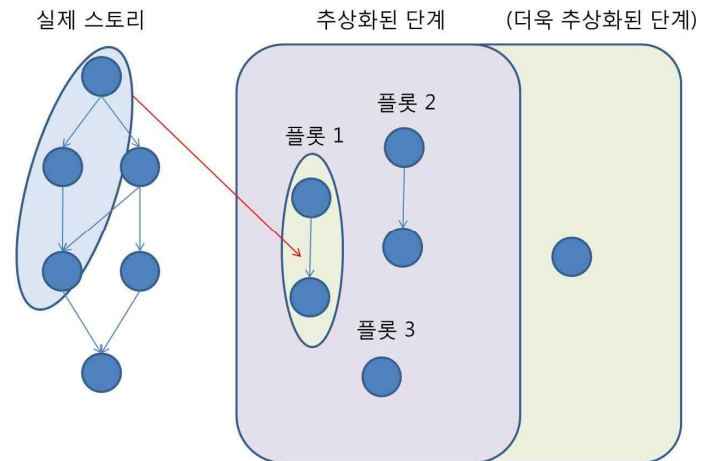


그림 1. 스토리 추상화 과정

이러한 스토리 및 구조를 파악함에 있어서 본 논문에서는 동화 속 등장인물을 기준으로 하여 스토리를 구성하는 사건들을 추출한 다음 등장인물별로 플롯을 구성한다. 이러한 방식을 택한 이유는 다음과 같다. 첫째로, 동화는 길이가 대체적으로 짧고 문장의 복잡도가 그다지

높지 않지만, 상대적으로 다양한 형태의 내용을 담고 있다. 둘째로, 등장인물을 기준으로 삼는 이유는 텍스트에서 스토리를 구성하는 사건들은 주로 해당 텍스트의 등장인물의 행위에 의하여 표현되기 때문이다.

본 논문의 구성은 다음과 같다. 2절에서는 관련 연구에 대하여 간략히 소개를 한다. 3절에서는 사건의 흐름을 분석하기 위하여 기본적으로 PMI기법을 적용하는 방식에 대하여 서술하고, 4절에서는 언어학적인 단서를 바탕으로 PMI 기법에서 나타날 수 있는 한계점을 보완하는 법을 논의한다. 5절에서는 제시된 방법으로 실험을 한 결과에 대한 평가 결과를 제시하며, 마지막으로 6절에서는 결론 및 향후 계획에 대하여 논의한다.

## 2. 관련연구

동화의 스토리를 이해하기 위하여 구조를 분석하는 연구가 간헐적으로 진행 중에 있다. 대표적으로 Propp은 러시아 민담 중에서 마법담 100개에 대한 분석을 통하여 민담에서 나타나는 대표적인 31개의 구조 유형을 제시하였다[1]. Propp이 분석한 구조 31가지 유형은 각각 일정한 순서가 정해져 있어 순차적인 구조를 가지며 몇 개의 구조의 조합을 통하여 민담이 구성된다고 하였다. 그러나 이러한 방법은 러시아 민담 중 마법담만을 대상으로 분석을 하였고, Propp이 분류한 31가지 세부구조를 순차적으로 조합하여 분석을 하였기에 다양한 내용의 텍스트에 대해 적용하기가 어렵다는 한계를 지닌다.

이에 반해, 자연어 처리 분야에서 동화의 세부 구조를 정의하고 상황에 맞게 적용하는 연구들이 있다. Goyal은 Lehnert가 정의한 플롯 유닛(Plot Unit)[2]을 동화 텍스트에 적용하였다[3]. 플롯 유닛은 등장인물의 감정 상태 변화에 따라 정의된 구조로, 이 연구에서는 감정을 나타내는 단어들을 수집하고, 이를 통해 문장 내의 감정 분석을 통해 등장인물의 감정 변화를 파악한 뒤 이미 정해진 플롯 유닛에 적용되도록 하였다. 하지만 이러한 연구는 문장 내에서 등장인물이 일으키는 사건에 초점을 맞추기보다 문장 내의 인물의 감정에 초점을 맞추었기에 본 논문에서 제안하는 연구와는 차이가 있다.

또한 Falconer는 동화 속에서 일어나는 내용을 세분화하여 임의의 단위로 구분 한 뒤, 등장인물과 벌어지는 사건을 추출하여 적용시키는 연구를 진행하였다. 하지만 이 연구는 자체적으로 사건들의 집합으로 구성된 단위 체계를 정의하였다는 한계를 지닌다.

비록 동화는 아니지만 다른 도메인에서도 스토리 내 사건의 흐름을 분석하는 연구가 진행되고 있다. 그 중

Chambers와 Jurafsky의 연구에서는 뉴스 기사에서 일어나는 사건을 추출하고 이를 연결하는 연구를 진행하였다[5]. 이 연구에서는 PMI(Pointwise Mutual Information) 기법을 사용하여 두 개의 사건이 같이 일어날 수 있는 확률을 계산하고 이를 통해 일련의 사건의 흐름에서 무작위로 하나의 사건을 제거한 뒤 빈 곳에 해당될 수 있는 적절한 사건을 추출하는 연구를 진행하였다. 본 연구는 그들의 연구 방법을 토대로 기계학습 기법을 이용한 결과에서 나타날 수 있는 한계를 보완하기 위하여 언어학적인 요소를 분석하여 적용하는 것에 초점을 맞춘다.

## 3. PMI를 이용한 사건의 흐름 분석

텍스트 내에서 일어나는 사건의 흐름을 분석하는 Chambers와 Jurafsky의 연구에서는 PMI를 이용하였다. PMI는 두 개의 사건이 동시 발생할 수 있는 확률을 구하는 대표적인 방법이다. 이를 텍스트 내에서 나타나는 사건의 흐름을 파악하는데 적용하기 위하여 동일한 등장인물이 속하는 사건들을 추출하고 이를 통해 PMI를 계산한다. 이 중 사건은 보통 등장인물의 직접적인 행위를 통해 벌어지므로, 서술어(동사+목적어)를 분석하는 것이 적합하다. 구체적인 행위를 판단하기 위해 목적어 등 까지 적용하는 것은 복잡도가 커지며, 표본이 적어지기때문에 확률적으로도 변별력 있는 수치를 구하기 힘들다. 그러므로 등장인물이 어떠한 행동을 하고 당하는지에 해당하는 흐름을 파악하는 것에 초점을 맞추기 때문에 동사만 분석하는 방법을 취한다. PMI를 구하기 위한 식은 다음과 같다[5].

$$pmi(e(w, d), e(v, g)) = \log \frac{P(e(w, d), e(v, g))}{P(e(w, d))P(e(v, g))}$$

$$P(e(w, d), e(v, g)) = \frac{C(e(w, d), e(v, g))}{\sum_{x,y} \sum_{d,f} C(e(x, d), e(y, f))}$$

위의 식에서 사건  $e(w,d)$ 는 동사와 의존(dependent)의 쌍(pair)을 의미하며  $C(e(x,d), e(y,f))$ 는 동일 대상이 속하는 두 개의 사건  $e(x,d)$ 와  $e(y,f)$ 의 개수를 의미한다.

영어와는 달리 한국어는 어미와 어간의 결합, 그리고 조사의 사용 등으로 인하여 동사라 할지라도 그 변형이 매우 다양하다. 그러므로 본 연구에서는 다음과 같은 방

법으로 처리하였다. 본 연구에서는 형태소 분석을 통하여 품사를 추출하고 이 중 동사와 행위성 명사로 분석된 단어가 사건을 나타낸다고 판단한다. 또한 한국어에서는 자동사와 피동사가 각각 존재한다. 특히 어근의 의미는 같더라도 단순히 형태소 분석만으로는 두 개의 동사가 다르게 분석될 수 있기에 기본적으로 분석 결과를 자동사의 형태로 변형을 한 뒤 PMI를 계산한다. 또한 의존 부분은 사건을 행하는 동작주(Agent), 그리고 사건을 당하는 대상주(Patient)로 구분한다. 이렇게 계산된 PMI를 바탕으로 제일 높은 값을 가지는 사건들을 결합하여 체인을 구성한다. 이를 통해 어떠한 사건 뒤에 주로 등장하는 사건을 확률적으로 계산하여 선별할 수 있으며 해당 사건들을 나열함으로써 스토리를 단순화 할 수 있다.

#### 4. 언어학적 분석을 통한 사건의 흐름 분석

위와 같이 PMI를 이용하면 등장인물이 행하거나 당하는 주요한 사건들을 얻을 수 있기는 하지만 이러한 방식

사건의 시간적 흐름과 사건의 목적을 아는 것이 중요하다. 시간에 따른 사건의 변화는 비슷한 사건들의 흐름에 이어 새로운 사건이 발생하는 가능성을 내포하기 때문이다. 또한 사건은 기본적으로 등장인물의 행위를 통하여 발생하는데, 인물이 특정한 목적을 가지고 행한 사건은 등장인물의 의도가 담겨있다고 볼 수 있기에 스토리 이해에 있어서 중요하다고 볼 수 있다. 이러한 두 가지 요소들을 분석하는 것은 단순히 통사적인 부분뿐만 아니라 의미적인 부분까지도 파악해야 하므로 좀 더 복잡한 분석 기법을 요구한다. 이러한 고려 사항을 배경으로 본 연구에서는 형태소 분석 결과를 기반으로 한국어 문법 구성 요소 중 시간과 목적을 나타내는 부사절과 그 외의 언어학적 단서를 통해 스토리 이해에 있어 필요한 사건을 추가하도록 하였다.

##### 4.1 부사절 사용을 통한 단서

###### 4.1.1. 시간의 부사절

한국어에서는 부사형 어미들을 통하여 시간적으로 선행절과 후행절을 구분할 수 있다[6]. 이 중 어떠한 부사형 어미들을 사용하는지에 따라서 두 가지의 사건이 동시적으로 발생한다고 볼 수 있는 것과 계기적으로 이어짐을 보여 주는 것으로 분류할 수 있다.

(1) (a) 늑대는 **울며** 숲속으로 **달아났어요**.

(b) 늑대는 문을 **열자마자** 양들에게 **달려들었어요**.

(1a)의 경우, 선행절과 후행절 사이의 ‘-며’란 어미가 쓰였다. ‘-으며/며’는 동시적 관계를 나타내주는 대표적인 어미로 이를 바탕으로 볼 때 두 사건간의 시간적 차이가 없다는 것을 알 수 있다. 이 경우에는 두 가지 사건을 모두 취한다. (1b)와 같이 ‘-자마자’는 선행절과 후행절의 시간관계를 구분할 수 있지만, 의미적으로 시간적 차이가 거의 없다고 볼 수 있다. 그러므로 ‘-자/자마자’와 같이 선행절과 후행절의 시간적 차이가 존재하더라도 두 가지 사건을 모두 취한다.

(2) (a) 늑대는 아기 양 여섯 마리를 **찾아서** 통째로 **삼켜버렸어요**.

(b) 어느 날, 나무꾼은 연못가에서 나무를 **하다가** 도끼를 **빠뜨렸어요**.

예시 (1)과 같이 시간을 나타내는 부사형 어미 중 동

등장인물 : 백설 공주

현재 사건의 체인

(주다, 대상주) - (먹다, 동작주)

다음 사건 후보군

(깨다, 동작주) - 3.94

(쓰러지다, 동작주) - 2.94

(살다, 동작주) - 1.35

은 빈도수에 기반한 확률 계산을 사용하기 때문에 다음과 같은 문제점을 야기할 수 있다.

그림 2. PMI를 통한 사건 추출의 한 예

그림 2와 같이 백설 공주에서 백설 공주에 대한 플롯을 생성하는 과정에서 PMI를 통해 계산된 결과에 따르면 다음에 올 수 있는 사건의 후보군 중 우선적으로 선택 되는 것은 ‘깨다’라는 동사이다. 이 경우 사건의 흐름은 ‘주다 - 먹다 - 깨다’의 형태를 띠게 된다. 하지만 스토리의 이해도 측면에서는 ‘깨다’ 보다는 ‘쓰러지다’가 선택되는 것이 더욱 자연스럽다. 이러한 단점 부분을 보완하기 위하여 본 연구에서는 텍스트 자체에서 나타나는 언어학적 단서를 분석하였고, 이를 통해 추출되어야 하는 사건을 추가하였다.

스토리 이해를 위한 사건의 흐름을 분석함에 있어서

시성을 띠는 것과 달리 (2)의 예문들에서는 계기적인 성격을 띠는 어미가 사용되었다. (2a)의 예문에서는 부사형 어미인 ‘-아서’를 사용하여 선행절의 동사인 ‘찾다’와 ‘삼키다’ 사이의 시간의 차이가 명확히 있음을 나타낸다. 또한 ‘-아서/어서’와 같은 어미의 사용은 시간의 차이를 나타내기도 하지만 선행절의 동사의 의미를 후행절의 원인 또는 수단으로 여기게끔 하기도 한다. 실제로 어떠한 사건들이 벌어졌는가에 초점을 맞출 경우 원인으로서는 발생하는 사건보다는 결과적으로 발생하는 사건이 더 중요하다. 따라서 (2a)와 같은 예의 경우에서는 후행절의 동사만 취한다. (2b)의 예문에서는 ‘-다가’라는 어미가 쓰였다. ‘-다가’는 동사와 결합하여 선행절의 사태가 중단되고 후행절에서 그와 대립적인 사태가 일어남을 보이는 어미로서 쓰인다. 이를 통해 시간의 변화가 있음을 알 수 있을 뿐만 아니라 의미적으로도 비슷한 사건이 발생하다가 그와 반대되는 새로운 사건이 발생한다는 것을 알 수 있다. 그러므로 후행절의 동사에 초점을 맞추어 사건을 구성하는 동사로 후행절의 동사를 취한다.

#### 4.1.2. 목적의 부사절

한국어의 부사형 어미 중에는 목적을 나타내는 어미들이 있다. 이들은 주로 ‘하기 위해’라는 의미를 지니는 어미들로서, 이러한 어미가 나타난 절은 목적을 나타내는 절로 실제로는 사건이 아직 발현되지 않음을 나타내지만, 이를 통해서 후행절의 사건이 결과로 간주됨을 알 수 있다.

(3) (a) 도둑은 **도망치려고** 담을 **넘고** 있었다.

(b) 그렇게 말하고서 왕비는 사냥꾼을 **불러** 그자리에서 **죽여버렸다**.

(3)의 예문에서처럼 ‘-으려고/려고’와 ‘-으러/러’ 같은 어미들의 사용은 후행절의 사건이 선행절에서의 사건을 목적으로 행해진다는 것을 나타낸다. 본 연구에서는 최종적으로 일어나는 사건에 초점을 맞추어 목적을 나타내는 선행절 대신 결과로 행해진 후행절의 사건을 취한다. 이러한 목적을 나타내는 어미로서 ‘-고자’도 비슷한 의미를 가진다. 이러한 어미들은 복문이 아닌 단문에서는 보조동사 ‘하다’와 어울려 쓰이는 경향이 있다. 이러한 경우에는 결과를 의미하는 후행절이 없으므로, 목적의 용도로 쓰인 동사를 취한다.

(4) (a) 한스는 그사이 **몰라보게** 변했어요.

(b) 아이들이 마음 놓고 놀 수 **있게** 놀이터를 크게

넓혔다.

‘-게’와 같은 경우도 목적의 의미를 지닌 어미로 사용되는 경우가 있다. 하지만 ‘-게’는 (4a)의 예시처럼 부사어에 쓰이는 경우가 많으며, 이때는 어떠한 목적을 내포한다는 의미를 갖지 않는다. 이러한 분석은 문장의 의미를 제대로 이해해야만 정확히 판단할 수 있는 문제이기에 일반성이 있게 처리하는 것은 간단하지 않다. 그러므로 본 연구에서는 (4b)에서와 같이 보조동사 ‘있다’와 결합한 ‘있게’의 경우만을 분석 대상으로 삼는다. 이 경우에는 선행절이 목적을 의미하는 것이 명확해지므로, 앞에서의 접근 방법과 같이 후행절의 동사를 취한다.

#### 4.2 단어의 의미를 통한 단서

비록 형태소 분석을 통한 어미의 사용 방법을 통한 접근은 아니지만, 단어 자체가 의미하는 내용을 바탕으로 주어진 사건을 판단할 수 있다.

동화에서는 극적인 효과를 위하여 스토리의 흐름을 반전시키는 표현들이 나타난다는 특징이 있다.

(5) (a) 어느 날 양치기 소년은 장난을 치고 싶어졌어요.

(b) 그 때 숲속에 사는 늑대가 숨어서 이 모습을 지켜보고 있었어요.

(c) 신령님은 다시 물속으로 사라졌어요.

(5a)와 (5b)의 예시에서 나타난 어휘들은 특정한 시간을 묘사하여 이전의 분위기와는 전혀 다른 사건을 발생시켜서 분위기를 고조시키는 역할을 하게 한다. 이를 형태소 분석 결과로 나타내면 ‘지시관형사 + 명사’의 형태로 나타나게 되는데 이러한 결과는 다양하게 나타날 수 있다. 그렇기에 단어를 따로 설정해두기보다는 위와 같은 형태소 분석 결과를 얻었을 때 명사 부분의 의미를 참조한다. 이는 세종 사전을 참고한다. 세종 사전에서 의미 부류가 시간을 나타내는 명사와 결합하였을 경우에는 해당 절에서 발생하는 동사를 사건으로 추가적으로 취한다.

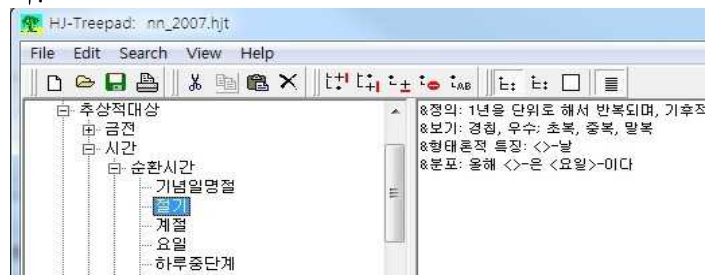


그림 3. 세종 사전에서 의미 부류의 계층구조

또한 (5c)와 같이 ‘다시’란 표현은 반복의 속성을 통해 이미 일어났던 사건을 강조하는 역할을 한다. 이 역시 형태소 분석을 통하면 부사로 분석된다. 위에서 기술한 방법과 동일하게 세종 사전을 참고하면, 의미 표시에서 ‘반복상’이라는 속성을 가진다는 것을 알 수 있다. 이를 통해 해당 절에서 발생하는 동사를 사건으로 추가로 취하도록 한다.

### 5. 실험 및 평가

실험을 위해서 25편의 동화에서 50개의 등장인물들이 행하는 사건들을 추출하여 학습 데이터로 삼아 PMI를 계산하였다. 각 동화는 형태소 분석을 통하여 사건으로 분류할 수 있는 동사와 행위성 명사를 추출하였고, 패턴을 설정하여 등장인물이 해당 사건의 동작주인지 대상주인지를 설정하였다. 동화는 보통 성인을 대상으로 한 작품에 대한 이해력이 떨어지는 유아들을 대상으로 하는 문학이므로 지시 대명사를 쓰기보다는 직접적으로 등장인물을 호칭하기에 참조 현상(coreference)의 해결은 고려할 부분이 적지만, 그렇지 않은 경우에 대해서는 수동으로 참조 현상을 해결하였다.

평가를 위해서는 학습에 쓰이지 않은 동화에서 24개의 등장인물이 행하는 사건들을 추출하였다. 평가를 위한 실험은 두 가지 형태로 진행되었다. 이미 학습된 PMI를 사용하여 테스트에 쓰일 각 등장인물의 사건의 흐름 중 현재 사건 다음에 올 수 있는 사건들을 선택하게 하므로 스토리를 단순화하도록 하였고 이를 baseline으로 삼았다. 또한 언어학적 단서를 통해 추가로 선택된 사건들을 baseline으로 뽑힌 사건들과 결합하였다. 이 방법을 통해 총 48개(24\*2)의 단순화 된 사건을 얻을 수 있었다.

평가는 총 5명의 사용자 평가를 통해 이루어졌다. 사용자 평가는 리카트 5점 척도를 사용하였고, 5점이 가장 높은 자연스러움을 나타내는 척도로서 설정하였다. 평가 항목은 크게 두 가지 형태로 진행되었다. 첫 번째 평가 항목은 동화의 스토리를 알고 있을 경우, 각 캐릭터마다 실험을 통해 추출된 단순화된 사건들의 집합이 실제 동화의 스토리를 이해하는데 얼마나 자연스러운가를 평가하였다. 두 번째 평가 항목은 동화의 내용을 모른다고 가정하고 각 캐릭터의 사건들의 집합의 흐름만 보았을 때, 어떠한 인물이 행동하는 과정이란 측면에서 얼마나 자연스러운가를 평가하였다. 사용자들의 평가를 분석한 결과는 다음과 같다.

표1. 첫 번째 평가의 결과

|     | B<br>평균 | B<br>표준편차 | B+L<br>평균 | B+L<br>표준편차 |
|-----|---------|-----------|-----------|-------------|
| U1  | 2.0417  | 0.8065    | 3.7917    | 0.9315      |
| U2  | 2.6667  | 0.9168    | 3.8333    | 1.007       |
| U3  | 2.5417  | 0.833     | 4.0833    | 0.7173      |
| U4  | 1.9583  | 0.6903    | 3.625     | 0.9237      |
| U5  | 2.0417  | 0.8065    | 3.8333    | 1.0072      |
| All | 2.25    | 0.8526    | 3.8333    | 0.9198      |

표2. 두 번째 평가의 결과

|     | B<br>평균 | B<br>표준편차 | B+L<br>평균 | B+L<br>표준편차 |
|-----|---------|-----------|-----------|-------------|
| U1  | 3.0417  | 0.6903    | 3.2083    | 0.5882      |
| U2  | 2.625   | 1.0555    | 3.5       | 0.8341      |
| U3  | 2.2083  | 0.8836    | 3.75      | 0.794       |
| U4  | 3       | 0.7223    | 3.16667   | 0.702       |
| U5  | 2.2917  | 0.8065    | 3.0417    | 0.55        |
| All | 2.6333  | 0.8976    | 3.3333    | 0.7372      |

표1과 표2는 두 가지 평가에서 분석된 결과를 나타내고 있다. U1에서 U5까지는 응시자들이며, B는 baseline, B+L은 baseline인 PMI 기법과 언어학적 분석을 통한 결과에 대한 평가를 나타낸다. 두 가지 평가에 대하여 ANOVA 테스트를 통해 baseline과 추가적인 방법을 사용한 실험사이의 점수의 차이가 있음을 확인할 수 있다. (significant (p < 0.003) )

첫 번째 평가의 결과를 보면 언어학적 단서를 추가하여 등장인물이 행한 사건을 분석한 것이 스토리의 이해도 면에서 더욱 자연스러움을 알 수 있다. 하지만 두 번째 평가의 결과를 보면 비록 언어학적 단서를 추가한 것이 PMI만을 통하여 분석한 결과에 비하여 좀 더 자연스럽기는 하지만 첫 번째 평가보다 점수가 낮아 떨어진 것을 확인할 수 있다. 또한 baseline의 점수도 첫 번째 평가보다 상승한 것을 알 수 있다. 이는 스토리와는 무관

하게 어떠한 인물이 행동하는 일련의 사건들이라는 관점에서 보면 추출되어진 사건의 흐름이 자연스럽지 않다는 것을 의미한다. 이는 등장인물이 사건을 행하는 목적이거나 내용에 대한 배경 지식이 없는 상태이므로, 시간의 흐름에 따라서 어떠한 행동을 하려하는 것인지 파악하는 것이 힘들기 때문이다. 더불어 평가에 사용된 실험 결과는 어떠한 시점 혹은 상황 속에서의 사건들만을 추출한 것이 아닌 동화 내용 전체에서 사건들을 추출한 것이다. 결국 단순화 과정에서 생략되는 정보들로 인하여 갑자기 서로 관계가 없을 듯한 사건들을 접하게 되므로 자연스럽지 않을 수 있다. 이러한 부분은 동일 등장인물의 사건의 집합 중에서도 더욱 세분화 된 플롯을 만들어 각각이 의미하는 바를 분석하여 묶을 수 있다면 더욱 자연스러울 것이다. 이는 향후 연구에서 해결되어야 할 부분으로 남긴다.

## 6. 결론

본 논문에서는 동화에서 사건의 흐름을 추상화하는 것을 목적으로 하여, 그를 위한 선행 연구로 사건의 흐름을 파악하는 방법에 대하여 제안하였다. 이를 위해 동화 속의 등장인물들을 바탕으로 각 등장인물들이 행하거나 당하는 사건들을 파악하기 위해 PMI 기법을 사용하여 기본적으로 해당 캐릭터가 행할 수 있는 사건들을 판단하였다. 또한 스토리 이해에 있어서 중요한 사건이지만 미처 추출되지 않은 사건들을 추가하기 위하여 한국어 문법 요소 중 부사절의 사용을 단서로 삼았다. 시간의 부사절과 목적의 부사절을 찾기 위해 형태소분석을 통하여 이러한 부사절을 형성하는 부사절어미를 파악하였다. 이를 통하여 추가해야할 사건을 찾고 기본적인 방법에서 드러날 수 있는 한계를 보완하는 방법을 제시하였다. 이러한 방법은 기계학습을 통해 사건의 흐름을 분석하는 연구에도 도움을 줄 수 있을 것이다.

그렇지만 평가에서 나타난 것과 같이 보다 자연스러운 사건의 흐름을 파악하는 것은 후속 연구가 필요하다. 또한 이러한 공통된 사건을 바탕으로 추상화 단계를 구축하는 것도 궁극적으로 연구되어야 할 한다. 특히 추상화 단계를 설정하는 문제에 있어서 어느 정도의 범위의 사건의 흐름을 한 개의 단위로 묶는지를 임의로 정하는 것은 정당성 문제를 야기할 수 있다. 그러므로 향후 연구로는 등장 인물간의 플롯의 유사도를 비교하여 적절한 범위를 찾아 플롯에 속한 사건들을 분리하는 연구를 진행할 예정이다. 그렇게 제외된 세부 플롯을 다시 단순화 시킴으로 한 단계 더 추상화 된 사건을 묘사할 수 있을

것이라 예상한다.

또한 등장인물을 기반으로 사건의 흐름을 파악한 것의 역으로 유사한 사건의 흐름을 가지는 등장인물들을 묶을 수 있다. 이를 통해 등장인물의 역할이 무엇인지를 좀 더 정확히 파악할 수 있을 것이다. 이 또한 향후 연구에서 진행되어야 할 부분이다.

## 감사의 글

이 논문은 지식경제부 지원으로 수행하는 21세기 프론티어 연구개발사업(인간기능 생활지원 지능로봇 기술개발사업)의 일환으로 수행되었다. 그리고 본 연구를 진행하는데 도움을 주신 민혜진님에게 깊이 감사드립니다.

## 참고문헌

- [1] V. J. Propp, Morphology of the Folktale, University of Texas Press, Austin, 1968.
- [2] W. Lehnert, Plot Units and Narrative Summarization, Cognitive Science, 5(4):293-331, 1981.
- [3] Amit Goyal et al, Automatically Producing Plot Unit Representations for Narrative Text, Proceedings of the 2010 Conference on EMNLP, 2010.
- [4] Sean M. Falconer et al, ETS Representation of Fairy Tales, Proceedings of ICPR 2004 Satellite Workshop, "Pattern representation and the future of pattern recognition: A program for action", Cambridge, UK, August 2004.
- [5] N. Cahmbers and D. Jurafsky, Unsupervised learning of narrative event chains, Proceedings of ACL-08:HLT, Hawaii, USA, 2008.
- [6] 이익섭, 한국어 문법, SNU Press, 2005