

텍스트 내 사건-공간 표현 간 참조 관계 분석을 위한 말뭉치 주석

정진우 이희진 박종철*†

한국과학기술원

Jin-Woo Chung, Hee-Jin Lee and Jong C. Park. 2014. Corpus Annotation for the Linguistic Analysis of Reference Relations between Event and Spatial Expressions in Text. *Language and Information* 18.2, 141–168. Recognizing spatial information associated with events expressed in natural language text is essential not only for the interpretation of such events and but also for the understanding of the relations among them. However, spatial information is rarely mentioned as compared to events and the association between event and spatial expressions is also highly implicit in a text. This would make it difficult to automate the extraction of spatial information associated with events from the text. In this paper, we give a linguistic analysis of how spatial expressions are associated with event expressions in a text. We first present issues in annotating narrative texts with reference relations between event and spatial expressions, and then discuss surface-level linguistic characteristics of such relations based on the annotated corpus to give a helpful insight into developing an automated recognition method. (Korea Advanced Institute of Science and Technology)

Key words: event, event expression, space, spatial expression, reference relation, corpus annotation

1. 서론

사건은 일반적으로 그것이 발생한 공간을 수반한다. 즉, 사건을 행위나 상태 변화와 같이 무엇인가 ‘발생하는 것’으로 규정할 경우, 그것이 발생할 당시의 공간적 범위가 상정되어 있다고 볼 수 있다. 또한 주어진 사건에 대하여 그것이 어떤 공간에서 발생하였는지에 대한 정보는 해당 사건 자체의 의미 해석뿐만 아니라 다른 사건과의 관계 해석에 있어서도 중요한 요소이다. 이러한 공간 정보는 시간 정보와 더불어 정보 추출, 질의 응답과 같은 자연언어처리의 주요 응용 분야에 유용하게 활용될 수

* 305-701, 대전광역시 유성구 대학로 291, 한국과학기술원 전산학과, E-mail: jwchung@nlp.kaist.ac.kr, heejin@nlp.kaist.ac.kr, park@nlp.kaist.ac.kr (교신저자)

† 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 콘텐츠산업기술지원사업으로 수행되었음.

있다. 예를 들어 “2013년에 교통 사고가 가장 많이 난 곳은?”과 같은 질문에 대답하기 위해서는 ‘교통 사고’라는 사건이 발생한 공간에 대한 정보를 ‘2013년’이라는 시간 정보와 연결하여 관련 정보를 추출할 수 있어야 한다.

사건과 관련한 공간 정보는 텍스트상에 명시적으로 나타나있는 경우도 있으나 그렇지 않은 경우도 있다. 아래의 문장들은 이에 대한 예시를 보인다.

- (1) 어젯밤 인천의 한 물류창고에서 화재가 발생했다.
- (2) 북한의 단속정이 백령도 북방한계선을 침범하여 우리 군은 6발의 경고사격을 실시했다.
- (3) 방망이를 잃어버린 도깨비는 터덜터덜 마을로 내려왔어요.

예문 (1)의 경우, ‘화재가 발생했다’라는 사건 표현과 ‘인천의 한 물류창고’라는 공간 표현이 한 문장 내에서 처소격조사라는 통사적인 장치를 통해 긴밀하게 연결되어 있기 때문에 사건에 대한 공간 정보를 비교적 쉽게 파악할 수 있다. 반면, 예문 (2)에서 ‘경고사격 실시’가 ‘북방한계선’ 부근에서 발생했다는 것을 파악하려면 형태-통사적인 분석뿐만 아니라 사건에 대한 의미 해석 및 사건 간 관계 추론이 필요하다. 또한 예문 (3)에서 ‘방망이를 잃어버린’ 장소는 해당 문장에 아예 기술이 되어 있지 않기 때문에 이를 파악하려면 문장을 넘어선 담화 단위의 분석이 필요하다. 위와 같이 사건에 대한 공간 정보가 텍스트상에서 형태-통사적으로 긴밀하게 연결되어 있지 않고 암시적으로 나타나는 경우 주어진 사건 표현에 대해서 그것이 발생한 곳을 지칭하는 공간 표현을 찾는 것은 쉽지 않은 문제이다. 또한 일반적으로 텍스트에서는 다수의 공간 표현이 함께 언급되기 때문에, 사건 표현에 해당하는 공간 표현을 선택하는 과정에서 애매성이 발생할 수 있다.

이러한 문제점을 고려하여 본 연구에서는 텍스트상에서 사건 표현과 공간 표현이 어떤 언어적 장치를 통해 연결되어 있는지를 계산학적으로 분석하는 것에 초점을 두었으며, 구체적으로 아래 질문에 대한 답을 구하는 것을 목표로 하였다.

- (4) a. 특정 사건 표현에 대하여 텍스트에 언급된 여러 공간 표현 중 하나를 선택하는 과정에서 애매성은 없는가? 즉, 주어진 사건이 발생한 공간을 지칭하는 표현을 텍스트상에서 하나로 결정하기 어려운 경우가 있는가?
- b. 특정 사건 표현에 대하여 하나의 공간 표현을 선택할 수 있다면, 그것을 선택하는 근거는 무엇인가?
- c. 주어진 사건 표현과 그것에 대하여 선택된 공간 표현이 텍스트상에서 어떤 표층적 관계를 맺고 있는가?

위 질문에 대한 답을 구하기 위해 본 연구에서는 사건의 흐름이 드러난 텍스트를 대상으로 두 명의 주석자가 사건 표현과 공간 표현을 주석하여, 주석 과정에서 발생한 문제들을 살펴보고 주석된 말뭉치에 나타나는 특징들을 분석하였다. 주석 말뭉치 구축을 위하여 본 논문에서는 먼저 사건의 정의와 사건 표현의 판별 기준을

제시하고, 사건 표현이 주어졌을 때 해당 사건이 발생한 곳을 지칭하는 공간 표현을 선택하는 기준에 대하여 논의한다. 본 연구는 사건과 공간 정보를 텍스트 수준에서 분석하는 것에 초점을 두었으며, 텍스트에 드러나지 않은 정보를 텍스트 외의 외부 자원을 활용하여 분석하는 것은 본 연구의 범위를 벗어나는 것으로 본다. 본 연구에서는 사건의 흐름이 뚜렷하면서도 사건 표현 대비 공간 표현이 드물게 언급되는 동화 텍스트를 분석 대상으로 삼았다.

본 논문의 나머지는 다음과 같은 순서로 기술된다. 2절에서는 텍스트에서의 공간 정보 분석에 관한 연구에 대하여 살펴 본다. 3절에서는 사건 표현의 판별 및 주석 기준을 제시하고 4절에서는 사건 표현이 참조하는 공간 표현을 판별하는 기준에 대하여 논의한다. 5절에서는 사건-공간 표현 간 참조의 애매성을 유형별로 분석하기 위한 단계별 주석 방법에 대하여 논의한다. 6절에서는 3, 4, 5절에서 각각 제시한 기준을 바탕으로 주석 진행 방식에 대하여 설명하고 주석 결과를 기술한다. 7절에서는 주석 결과에 대한 분석을 제시하고 8절의 결론 및 향후 계획으로 본 논문을 맺는다.

2. 관련연구

2.1 사건 및 시간 관련 연구

사건의 속성 중 시간에 대한 연구는 공간에 대한 연구에 비하여 자연언어처리 분야에서 오래 전부터 시작되어 활발히 이루어져 왔다. 특히 자연언어로 표현된 사건 및 시간 정보의 명세화를 위해 개발된 마크업 언어인 TimeML 및 이를 바탕으로 텍스트를 주석하여 구축된 TimeBank가 발표된 이후 (Pustejovsky *et al.*, 2003a; Pustejovsky *et al.*, 2003b), 자연언어처리 기술을 응용한 다양한 시간 정보 추출 연구가 꾸준히 이루어져 왔다 (Chambers *et al.*, 2007; Lapata & Lascarides, 2006; Mani *et al.*, 2006; Yoshikawa *et al.*, 2009; Mirza & Tonelli, 2014). 국내에서도 한국어 텍스트에 대한 시간 정보 추출 연구가 시도된 바 있으며 (김평 외, 2003; 김평 & 맹성현, 2006; 정영미 & 김용광, 2008), 최근에는 TimeML 및 TimeBank를 한국어에 적용할 수 있도록 사건 및 시간 정보 주석 기준을 명세화하는 프로젝트가 진행된 바 있다 (임서현 외, 2009; 유현조 외, 2009; 유현조 외, 2011). 특히 TimeML을 한국어에 적용하는 연구는 텍스트상의 사건들을 판별하고 분석하는 논의를 포함하고 있어, 본 연구에서도 공간 정보 분석에 앞서 사건 표현 판별 기준의 명세화를 위해 해당 연구들을 참고하였다. 이와 관련된 논의는 3절에서 제시한다.

2.2 사건 및 공간 관련 연구

사건과 관련한 공간에 대한 연구는 국외에서도 최근에서야 본격적으로 이루어지기 시작했다. 먼저 TimeML과 유사한 목적으로 고안된 SpatialML이 있는데, 이는 텍스트에 드러난 장소 및 지명 표현을 그것이 지칭하는 실제 지리적 위치 및 좌표, 또는 지명 사전 (gazetteer)에 등재된 지리 항목에 대응시키기 위해 제안된 마크업 언어이다 (Mani *et al.*, 2008). 이 연구에서는 “Rome”, “Rochester, NY”, “southern Kerala

district of Cudallah”와 같이 특정 장소를 직접 참조하는 표현과 “thirty miles north of Boston”, “an underpass beneath Pushkin Square”, “in the vicinity of Georgetown University”와 같이 특정 장소에 대한 상대적 위치를 지칭하는 표현, 그리고 그들 간의 위상적 또는 방향적 관계를 지칭하는 “beneath”, “to the left”와 같은 표현들을 분석 대상으로 삼고 있으며 이를 주석하기 위한 가이드라인을 제시하고 있다. 다음 문장은 SpatialML의 구체적인 예시를 보여준다.

(5) A town some 50 miles south of Salzburg in the central Austrian Alps

위 예문의 경우 SpatialML에서는 ‘town’, ‘Salzburg’, ‘Austrian’, ‘Alps’를 지명 표현으로 주석하고, 그와 더불어 ‘town’이 ‘Salzburg’을 기준으로 남쪽 방향으로 50마일 떨어져 있다는 거리 및 방향 정보와 ‘Salzburg’가 ‘Alps’에 지리적으로 포함되어 있고 ‘Alps’가 ‘Austrian’에 지리적으로 포함되어 있다는 정보를 함께 주석한다. 이와 같이 SpatialML은 객체 간의 지리적 관계 분석에 초점이 맞춰져 있다는 점에서 사건-공간 표현 간의 참조 관계를 분석하는 본 연구와는 거리가 있다.

텍스트에서 객체 간의 공간 관계를 분석하는 또 다른 연구로 공간 의미역 분석 (Spatial Role Labeling) 연구를 들 수 있다 (Kordjamshidi *et al.*, 2010). 아래 문장은 이에 대한 예시를 보여준다.

(6) Give me the book on AI on the big table behind the wall.

위와 같은 문장이 주어졌을 때, 공간 의미역 분석에서는 객체 간 공간 관계를 암시하는 ‘on’, ‘behind’에 대하여 해당 관계에 참여하는 세 객체 표현 ‘book’, ‘table’, ‘wall’을 찾고 $\langle on_{SPATIAL_INDICATOR} book_{TRAJECTORY} table_{LANDMARK} \rangle$, $\langle behind_{SPATIAL_INDICATOR} table_{TRAJECTORY} wall_{LANDMARK} \rangle$ 와 같이 공간적 관계를 나타내는 형식 표현을 생성한다. 이와 관련하여 SemEval-2012 Task 3 (Kordjamshidi *et al.*, 2012) 대회가 개최되었으며, 이어서 두 번째로 개최된 SemEval-2013 Task 3 (Kolomiyets *et al.*, 2013)에서는 움직임 (motion)에 대한 경로, 방향, 거리를 주석 및 추출하는 작업이 새로 추가된 바 있다. 그리고 이와 유사한 연구로 시간과 공간의 측면을 함께 고려한 Spatiotemporal Markup Language (STML, Pustejovsky & Moszkowicz, 2008)가 있으며, 최근에는 SpatialML과 STML의 개념을 통합한 ISO-Space (Pustejovsky *et al.*, 2011a; Pustejovsky *et al.*, 2011b)가 제안되어 ISO-TimeML과 같이 이를 국제 표준으로 제정하려는 시도가 이루어지고 있다. 앞서 제시한 연구들은 대부분 사건보다는 객체 간 공간 관계가 텍스트에 명시적으로 드러나 있을 때 이에 대한 정보를 주석하고 추출하는 것에 초점이 맞춰져 있다. 이 중 ISO-Space가 사건 주석에 대한 논의를 일부 포함하고는 있으나, 사건이 주석되는 경우는 그것이 문장 내에서 공간 관계를 지칭하는 특정 표현에 의해 다른 객체 또는 사건과 공간적 관계를 맺고 있는 경우로만 한정하고 있다. 아래 문장은 이러한 예시를 나타낸다.

- (7) The new tropical depression was about 430 miles (690 kilometers) west of the southernmost Cape Verde Island.

위 문장에서 사건을 지칭하는 ‘tropical depression’은 동일 문장 내에서 공간 관계를 지칭하는 ‘430 miles’, ‘west’에 의해 ‘Island’와 공간적 관계를 맺고 있으므로, ISO-Space는 이를 해당 공간 관계에 참여한다고 간주하여 주석한다. 위 예문의 경우 ISO-TimeML에서는 ‘tropical depression’을 사건 표현으로 간주하나, ISO-Space에서는 ‘tropical depression’과 ‘Island’를 모두 공간 관계에 참여하는 공간 객체 (spatial entity)로 간주하여 두 표현을 서로 구분하지 않고 모두 ‘SPATIAL_ENTITY’라는 표지로 주석한다는 차이가 있다(Pustejovsky *et al.*, 2011a). 따라서 ISO-Space에서는 실제로 텍스트에 언급된 모든 사건을 주석하고 있지 않을 뿐만 아니라, 주석의 대상이 되는 사건도 실제로는 주석 과정에서 객체와 구별하지 않고 있다고 볼 수 있다.

지금까지 살펴 본 연구들은 주로 공간, 공간 관계 및 객체의 움직임을 지칭하는 표현이 텍스트에서 명시적으로 주어지진 경우에만 한정하여 이를 형식 언어로 정규화하는 작업에 초점이 맞춰져 있는 반면, 본 연구에서는 텍스트상의 모든 사건이 특정 공간을 상정한다고 가정하여, 각 사건 표현과 그것이 참조하는 공간 표현이 텍스트 상에서 서로 어떤 표층적인 관계를 맺고 있는지를 분석한다는 점에서 차이를 보인다.

사건의 공간성에 초점이 맞춰진 연구로 사건 표현 간의 공간적 포함 관계를 파악하고자 한 연구를 들 수 있다(Roberts *et al.*, 2012; Roberts *et al.*, 2013). 이 연구에서는 텍스트에 드러난 사건들이 서로 암시적으로 공간적 포함 관계를 맺고 있는 경우가 많다는 점에 주목하여 이에 관한 정보를 적절히 추출할 수 있을 경우 공간 추론 (spatial reasoning)에 도움이 될 수 있다고 보았다. 해당 연구에서 제시한 아래 예시는 두 사건의 공간적 포함 관계를 보여준다.

- (8) In October of 1985, four hijackers under his command took over the Italian cruise ship Achille Lauro and killed a wheelchair-bound American tourist, Leo Klinghoffer.

위 예문에서 ‘took’이 지칭하는 배를 탈취한 사건은 ‘cruise ship Achille Lauro’가 지칭하는 선박 전체에 걸쳐 발생한 반면에, ‘killed’가 지칭하는 관광객을 살해한 사건은 선박 내 희생자가 있는 공간의 주변에 한정되어 발생했다고 볼 수 있으므로 ‘took’ 사건이 ‘killed’ 사건을 공간적으로 포함한다고 간주하였다. 그리고 텍스트로부터 이러한 정보를 올바르게 추출할 경우 ‘killed’ 사건이 발생한 공간이 ‘cruise ship Achille Lauro’임을 추론할 수 있다고 보았다. 이 연구에서는 공간적 포함 관계 식별 문제를 관계 인식 (relation recognition) 및 관계 유형 분류 (relation type classification)의 두 단계로 나누어 접근하는 방법을 제안하였으며, 형태-통사적인 자질과 어휘의미론적인 자질에 기반한 감독 학습 (supervised learning) 방법을 적용하여 관계 인식과 관계 유형 분류에 대하여 각각 54%와 60%의 성능 (F1-score)

을 보고하였다. 이 연구는 사건에 대한 공간 정보를 분석한다는 점에서 본 연구와 공통점이 있으나, 사건 표현이 참조하는 공간 표현을 직접 판별하거나 두 표현 사이의 참조 관계를 분석하지는 않고 있으며, 주어진 두 사건의 공간적 포함 관계를 분류하는 데 있어서 그것이 참조하는 공간 표현이 주요 자질로 사용되지 않았다는 점에서 본 연구와 차이를 보인다.

3. 사건 표현의 주석 기준

본 절에서는 사건이 발생한 곳을 지칭하는 공간 표현의 주석에 앞서, 사건 표현을 판별하고 주석하는 방법에 대하여 논의한다.

본 연구에서는 사건을 ‘이야기 내 인물 또는 사물의 행위 및 상태변화’로, 사건 표현을 ‘사건을 지칭하기 위해 텍스트상에서 사용된 표현’으로 정의한다. 만약 어떤 표현이 둘 이상의 행위 및 상태변화를 지칭하거나, 또는 서로 다른 공간에서 발생할 수 있는 행위 및 상태변화를 지칭할 경우, 이를 더 작은 단위의 표현으로 나누어 각 사건 표현이 하나의 공간에서만 발생할 수 있는 하나의 행위 및 상태변화만을 지칭하도록 주석한다. 가령 “무릎을 꿇고 빌다”의 경우, ‘꿇다’와 ‘빌다’는 서로 다른 행위를 지칭하므로 별개의 사건 표현으로 주석한다. 또한 “용왕님을 만나려고 한다”의 경우, ‘만나다’가 지칭하는 물리적 행위와 ‘-려고 한다’가 지칭하는 심리적 행위는 서로 다른 공간에서 발생할 수 있으므로, ‘만나려고’와 ‘한다’를 별개의 사건 표현으로 주석한다. 이는 본 연구가 사건에 대한 공간 정보를 분석하는 데 초점이 맞춰져 있는 만큼, 서로 다른 공간에서 발생할 수 있는 행위 및 상태변화들은 최대한 별개로 분석하고자 하는 목적이 있기 때문이다. 그러나 본 연구가 사건에 대한 의미론적 해석 자체를 목적으로 하는 것은 아니므로, 사건 표현의 주석 기준을 세우는 데 있어 위에서 제시한 원칙을 따르되 비교적 명확한 형태론적 기준을 사용하여 주석의 애매성을 최소화하였다.¹ 본 연구에서는 아래에 나열된 기준을 통해 사건 표현을 판별한다.

동사 및 명사. 본 연구에서는 기본적으로 모든 동사와 행위성 명사를 사건으로 간주한다. 임서현 외 (2009), 유현조 외 (2009), 유현조 외 (2011)에서는 형용사 및 행위성 명사가 아닌 명사에 대하여 그것이 텍스트상에서 명백히 시간성을 갖는 경우에만 제한적으로 사건 표현으로 판별하였으나, 본 연구에서는 사건 표현의 공간성을 현재 단계에서 논하기에는 이르다고 판단하여 형용사 및 행위성 명사가 아닌

¹ 사건 표현의 주석 기준은 기본적으로 해당 주석이 어떠한 목적을 가지고 수행되는지에 따라 달라진다. 가령, 한국어 TimeML 연구 (유현조 외, 2011)에서는 텍스트 수준의 시간성 분석을 목적으로 하였기 때문에 일반적인 어휘론적 관점의 사건 판별 기준을 따르지 않고, 주어진 사건 표현이 텍스트 수준에서 시간성을 나타내고 있는 경우에만, 즉, 그것이 텍스트 내에서 시간 또는 시점을 지칭하는 특정 표현과 묶여있는 경우에만 사건으로 판별하였다. 본 연구에서도 이와 유사하게 공간성을 가진, 즉 텍스트에서 특정 공간 표현과 묶여있는 표현만을 공간 표현으로 판별하는 기준을 따를 수 있겠으나, 본 연구의 목적이 텍스트상에서 각 사건 표현이 특정 공간 표현과 묶여 있는지, 묶여있다면 어떤 방식으로 묶였는지를 살펴보는 것이므로, 본 연구에서는 우선 어휘론적인 기준을 도입하여 사건 표현을 판별하고, 각 사건 표현이 특정 공간 표현과 묶여있는지의 여부는 5절에서 사건-공간 표현 간 참조를 주석한 결과를 통해 별개로 논의하였다.

명사는 사건 표현으로 간주하지 않았다.

합성어. 합성어의 경우 하나의 사건을 지칭한다고 간주하여 합성어를 구성하는 각각의 형태소를 별개의 사건으로 주석하지 않는다. 가령 ‘뛰어놀다’라는 사건 표현의 경우 ‘뛰다’와 ‘놀다’로 분리하여 주석하지 않고 하나의 토큰으로 주석한다. 주어진 표현이 합성어인지의 여부는 표준국어대사전의 표제어로 등재되어 있는지 여부로 판단하도록 하였다.

명사 연쇄. 명사가 연쇄되어 명사구를 이룰 경우, 유현조 외(2009)에서와 마찬가지로 명사구의 핵어(head; 즉, 명사구의 마지막 단어)가 사건을 지칭하는 경우에만 명사구를 구성하는 각 명사들을 사건 표현의 후보로 간주한다. 가령, ‘경고 사격’의 경우 핵어인 ‘사격’이 행위를 지칭하므로 행위를 지칭하는 ‘경고’와 ‘사격’을 모두 주석한다.

기능동사. 한국어 텍스트에서는 영어와 마찬가지로 행위성 술어명사와 기능동사가 서로 통사적으로 연결되어 있으나 텍스트상에서 연속되어 있지 않은 경우가 빈번하게 있기 때문에 이에 대한 주석 원칙이 필요하다. 본 연구에서는 기능동사가 의미적 술어의 성격을 지니지 않는다고 판단하여 술어명사에만 주석하는 것을 원칙으로 하였으나, 기능동사에 대한 뚜렷한 판별 기준이 존재하지 않는 관계로 ‘술어명사+기능동사’의 연쇄가 표준국어대사전의 표제어로 존재하는 경우에만 기능동사를 별개로 주석하지 않고, 그 외의 경우에는 술어명사와 기능동사를 서로 다른 사건으로 간주하여 별개로 주석하는 기준을 따르도록 하였다. 이 경우 주석하지 않는 기능동사는 대부분 ‘-하다’ 또는 ‘-되다’에 해당하게 되나, 드물게 ‘도망을 치다’와 같이 일부 다른 기능동사들도 주석하지 않게 된다. 아래의 예시는 기능동사를 주석하는 경우와 주석하지 않는 경우를 나누어 보여준다.

(9) 기능 동사를 사건으로 주석하지 않는 예

‘결심을 하다’ (‘결심하다’가 표제어이므로 ‘결심’에만 주석)

‘걱정이 되다’ (‘걱정하다’가 표제어이므로 ‘걱정’에만 주석)

‘도망을 치다’ (‘도망치다’가 표제어이므로 ‘도망’에만 주석)

(10) 기능동사를 사건으로 주석하는 예

‘결심을 굳히다’ (‘결심굳히다’가 표제어가 아니므로 ‘결심’과 ‘굳히다’에 모두 주석)

‘걱정을 끼치다’ (‘걱정끼치다’가 표제어가 아니므로 ‘걱정’과 ‘끼치다’에 모두 주석)

‘도망을 다니다’ (‘도망다니다’가 표제어가 아니므로 ‘도망’과 ‘다니다’에 모두 주석)

본 연구의 주요 목적이 사건 표현이 참조하는 공간 표현에 대한 분석이므로, 기능동사를 별개의 사건으로 주석할지 여부는 해당 기능동사가 술어명사와 관계 없이 독립적으로 공간성을 내재하고 있는지 여부에 따라 결정하는 것이 바람직할 것으로 판단된다. 그러나 위에서 논의하였듯이 사건 주석 단계에서 기능동사의 공간성을

단정하기는 어려운 관계로 이와 같은 제한적인 기준을 사용하였다.

보조동사. 보조동사는 기본적으로 자립성이 없고 본용언에 문법적, 어휘적 의미를 추가해주는 역할만을 수행하므로 기능동사와 마찬가지로 의미적 술어의 성격을 갖지 않는다고 판단하여 이를 독립적인 사건으로 간주하지 않고 본용언과 연결하여 하나의 사건으로 간주한다. 가령 ‘집에 가고 있다’의 경우 ‘집에 가고 있다’와 같이 ‘가다’와 ‘-고 있다’를 묶어서 하나의 사건으로 간주하여 주석한다. 단, 예외적으로 양태를 나타내는 보조동사 중 아직 실현되지 않은 사건에 대한 화자의 태도나 판단을 나타내는 것들은 별개의 사건으로 간주한다. 이는 본 절의 도입부에서 논의하였듯이 해당 보조동사가 지칭하는 화자의 태도 및 판단이 실현된 공간과 본동사가 지칭하는 사건이 실현될 공간을 별개로 분석하기 위함이다. 가령, ‘집에 있으려고 했어요’와 같이 의도의 양태를 나타내는 예문의 경우, ‘집에 있다’라는 행위가 실현될 공간은 ‘집’인 반면에, ‘-려고 했다’라는 화자의 의도가 표출된 공간은 ‘집’이 아닐 수 있다는 가능성을 고려하여 ‘있으려고’와 ‘했다’를 별개의 사건으로 주석한다. 본 연구에서는 이러한 기준에 부합하는 아래의 보조동사들을 모두 본동사와 별개의 사건으로 간주하여 주석하며, 그 외의 모든 보조동사들은 주석하지 않도록 하였다. 유사한 역할을 하는 ‘-(ㄴ) 수 있다’(가능), ‘-고 싶다’(소망) 등의 보조형용사들의 경우 형용사는 사건으로 간주하지 않는다는 원칙에 따라 별개로 주석하지 않는다.

- (11) 사동: -게 하다 / -게 만들다 / -도록 하다
 의지/의도: -기로 하다 / -(으)려(고) 하다 / -고자 하다
 당위: -어야 하다 / -어야 되다

위에 제시한 원칙에 따라 “하늘나라에서 살 수 있게 해 주려고 했어요”와 같이 보조용언이 복합적으로 나타나는 문장은 다음과 같이 주석할 수 있다.

- (12) 하늘나라에서 살 수 있게 해 주려고 했어요.

위 예문에서 ‘살다’는 본동사, 두 ‘하다’는 위 목록에 열거된 보조동사이므로 사건으로 주석하며, ‘있다’는 보조형용사이므로 주석하지 않고, ‘주다’는 보조동사이나 위 목록에 열거되지 않았으므로 별개로 사건으로 간주하지 않고 본동사인 ‘하다’와 하나로 묶어서 주석한다.

접미사 및 선어말어미. 보조용언 외에 사동의 양태를 나타내는 접미사 ‘-시키다’와 ‘의지’의 양태를 나타내는 선어말어미 ‘-겠.’은 예외적으로 별개의 사건으로 간주한다.

보조용언, 접미사, 선어말어미와 같은 형태소 외에 특정한 의미론적, 화용론적인 장치를 통해서도 실현되지 않은 사건에 대한 양태가 나타날 수 있으나, 이를 판별하는 것은 본 연구의 범위를 벗어난다고 판단하여 위에서 열거한 형태론적인 기준만을 제한적으로 사용하였다. 또한 위에서 제시한 양태 범주 이외에 독립적인 공간성을 부여할 필요가 있는 또 다른 문법적 요소가 있는지에 대해서도 향후 추가 검토가 필요할 것으로 보인다.

이동 사건 표현. 위에서 제시한 기준에 따라 분류된 사건 표현 중, 그것이 ‘이동’의 의미를 내포하고 있을 경우 특별히 일반적인 사건 표현과는 별개로 ‘이동 사건 표현’으로 분류한다. 이동 사건 표현을 따로 분류하는 이유는 사건의 주체가 이동 행위를 하고 있을 경우, 그것의 출발점, 이동 경로, 도착점을 지칭하는 공간 표현을 모두 주석하려는 목적이 있기 때문이다. 단, 주어진 사건 표현을 이동 사건 표현으로 분류하는 경우는 세종 용언 상세사전상의 의미 부류가 [이동행위] 또는 그것의 하위부류에 해당하는 경우로만 한정한다. 이에 해당하는 표현으로 ‘가다’, ‘오다’, ‘해매다’, ‘달아나다’, ‘출발하다’, ‘찾아오다’ 등이 있다.

4. 공간 표현의 주석 기준

본 절에서는 3절에서 논의한 기준에 따라 주석된 각 사건 표현에 대하여, 그것이 지칭하는 사건이 발생한 공간을 지칭하는 표현의 판별 및 주석 기준에 대해 논의한다.

4.1 공간 표현의 정의

3절에서 논의한 것과 같이 본 연구에서는 공간을 사건과 연결된 개념으로 한정하고 사건을 ‘인물 및 사물의 행위 및 상태변화’로 정의하고 있으므로, **사건이 발생한 공간**을 ‘사건이 발생하는 기간에 걸쳐 이와 관련한 행위의 주체 또는 상태변화의 대상이 존재하는 공간’으로 정의한다. 또한 이에 따라 **사건 표현이 참조하는 공간 표현** 또는 **사건 표현에 대한 공간 표현**을 ‘사건 표현이 지칭하는 사건에 대하여, 그것이 발생한 공간을 지칭하는 표현’으로 정의한다. 따라서 사건-공간 표현 간 참조 여부를 판단하려면, 사건 표현이 지칭하는 사건에 대하여 그것이 발생한 시간적 범위 내에서 공간 표현이 지칭하는 물리적 공간에 사건의 주체 또는 대상이 존재하는지를 판단하는 것이 중요하다. 이를 위해서는 행위 및 상태변화에 대한 공간뿐만 아니라 그것이 지속된 시간을 함께 고려해야 한다.²

아래의 예시는 주어진 사건 표현에 대한 공간 표현을 찾는 과정을 보인다.

(13) 동생은 방망이를 들고 집으로 들어왔어. 방망이를 두드리자EVENT ...

위의 예문에서 ‘두드리자’라는 사건 표현의 경우, 그것이 지칭하는 행위가 지속되는 동안 사건의 주체인 ‘동생’이 존재한 공간은 문맥상 바로 이전 문장에 언급된 ‘집’이므로, 주어진 사건 표현 ‘두드리자’가 참조하는 공간 표현은 ‘집’이라고 볼 수 있다. 이와 같이 공간 표현은 항상 특정 사건 표현이 그것을 참조하는 경우에만 주석하도록 하며, 만약 특정 표현이 어휘의미론적으로 공간을 지칭하더라도 텍스트 내에서 어떤 사건 표현도 그것을 참조하지 않고 있다면 해당 표현을 공간 표현으로 주석하지 않는다.

² 이를 면밀하게 분석하기 위해서는 사건 및 공간 표현에 내재된 시공간적 속성들을 포상하여 그들간의 관계를 논리적으로 밝히는 과정이 필요할 것으로 보이나, 본 연구에서 다루는 주석 작업은 이러한 과정 없이도 주석자가 텍스트 수준의 분석만을 통해서 충분히 수행할 수 있다고 판단하여 본 연구에서는 이를 논의 대상에 포함시키지 않았다.

4.2 사건 표현에 대한 공간 표현의 판별 기준

주어진 사건에 대하여 특정 표현이 공간 표현인지를 판별하기 위해 아래의 기준을 사용한다.

- A. 주어진 사건 표현이 지칭하는 사건에 대하여, 그것이 발생한 공간을 가장 한정적으로 지칭하는 명사구를 해당 사건 표현에 대한 공간 표현으로 판별한다.

예) 강 건너 단양 고을_{SPACE} 원님은 그 소문을 듣고_{EVENT} ...

- ‘듣고’가 지칭하는 사건이 발생한 공간을 지칭하는 명사구로 ‘고을’, ‘단양 고을’, ‘강 건너 단양 고을’이 있으나, 이 중 해당 공간을 가장 한정적으로 지칭하는 명사구는 ‘강 건너 단양 고을’이므로, 이를 ‘듣고’에 대한 공간 표현으로 판별한다.

예) 강 건너에 사는 단양 고을_{SPACE} 원님은 그 소문을 듣고_{EVENT} ...

- ‘강 건너에 사는’이 ‘단양 고을’을 수식할 수 없으므로 (즉, ‘강 건너에 사는 단양 고을’이 독립적인 명사구가 될 수 없으므로) ‘듣고’에 대한 공간을 가장 한정적으로 지칭하는 명사구는 ‘단양 고을’이 되며, 이를 ‘듣고’에 대한 공간 표현으로 판별한다.

- B. 주어진 사건에 대하여 가장 좁은 공간을 지칭하는 표현을 그 사건에 대한 공간 표현으로 판별한다.

예) 취들은 재빨리 욱심쟁이 할머니 집의 안방_{SPACE} 으로 들어가_{EVENT} ...

- ‘들어가’가 지칭하는 사건에 대하여, ‘욕심쟁이 할머니 집’보다 ‘욕심쟁이 할머니 집의 안방’이 더 좁은 공간을 지칭하므로, 후자를 ‘들어가’에 대한 공간 표현으로 판별한다.

- C. 이동 사건 표현의 경우, 이동 경로를 아우르는 공간을 지칭하는 공간 표현, 이동 경로의 시작점을 지칭하는 공간 표현, 이동 경로의 도착점을 지칭하는 표현을 해당 사건에 대한 공간 표현으로 판별하여 각각을 SPACE-PATH, SPACE-SRC, SPACE-GOAL로 표지한다.

예) 그리고는 동생네 집_{SPACE-SRC} 에서 개암을 챙겨 산기슭_{SPACE-PATH} 을 따라 도깨비 집_{SPACE-GOAL} 으로 발걸음을 옮겼어요_{EVENT}.

- ‘산기슭’은 ‘옮겼어요’가 지칭하는 이동 행위의 경로를 아우르는 공간 표현, ‘집’은 출발점을 지칭하는 공간 표현, ‘도깨비 집’은 도착점을 지칭하는 공간 표현으로 판별한다.

D. 대명사 또는 지시형용사의 수식을 받는 명사구를 선택해야 할 경우, 그것의 선행사를 대신 선택한다.

예) 얼마 후, 인당수_{SPACE1} 에 커다랗고 고운 연꽃이 두둥실 떠올랐습니다.
 마침 그곳_{SPACE2} 을 지나던_{EVENT} 뱃사람들이 ...

- ‘지나던’에 대한 공간 표현으로 ‘그곳’ 대신에 그것의 선행사인 ‘인당수’를 선택한다.

E. 위의 기준으로 우선 순위를 구분할 수 없는 표현이 여러 개일 경우, 텍스트 상에서 사건 표현과 가장 가까운 것을 선택한다. 텍스트 상에서의 거리는 두 표현 사이에 존재하는 어절의 개수로 계산하며, 만약 주어진 사건 표현을 기준으로 앞과 뒤의 동일한 거리에 우선 순위를 구분할 수 없는 두 개의 공간 표현이 존재한다면, 앞에 언급된 것을 선택한다.

예) 형은 개암을 들고 산 속을 걷다가 허름한 도깨비 집_{SPACE1} 을 발견했어요. 그리고 집_{SPACE2} 에서 밤새 도깨비들을 기다렸어요_{EVENT}.

- ‘기다렸어요’가 지칭하는 사건이 발생한 공간을 지칭하는 표현으로 ‘허름한 도깨비 집’_{SPACE1} 과 ‘집’_{SPACE2} 이 있는데, 후자가 텍스트 상의 거리는 더 가까우나 전자가 해당 공간을 더 한정적으로 지칭하므로 전자를 ‘기다렸어요’에 대한 공간 표현으로 선택한다.

예) 얼른 강_{SPACE1} 을 건너기 시작했어요.
강_{SPACE2} 을 한참 건너는데_{EVENT} ...

- ‘건너는데’가 지칭하는 사건이 발생한 공간을 지칭하는 표현인 ‘강’_{SPACE1} 과 ‘강’_{SPACE2} 은 앞서 제시한 기준으로 우선순위를 구분할 수 없으므로, 텍스트 상의 거리가 더 가까운 후자를 선택한다.

F. 별개의 사건으로 주석된 보조용언, 접미사, 선어말어미의 경우 그것이 지칭하는 태도가 실현된 공간을 지칭하는 표현을 공간 표현으로 판별한다.

예) 늦게까지 산_{SPACE1} 에서 나무_{EVENT1} 를 하다가 집_{SPACE2} 에 가려고_{EVENT2} 하는데_{EVENT3} ...

- ‘집’은 ‘가려고’가 지칭하는 이동 행위의 도착점을 지칭하므로, ‘가려고’에 대한 공간 표현으로 선택한다.
- ‘산’은 ‘가려고 하다’라는 심리 행위가 실현된 공간이므로, ‘하는데’에 대한 공간 표현으로 선택한다.

G. 사건 표현을 구성하는 술어의 논항이 공간을 의미하더라도, 그것이 해당 사건이 발생한 공간을 지칭하지 않으면 공간 표현으로 판별하지 않는다.

예) 원님은 죄인을 고향으로 보냈어요EVENT.

- ‘고향’은 ‘보냈어요’의 논항에 해당하나, 실제 ‘보내다’라는 사건이 이루어진 공간(즉, ‘보내다’라는 사건이 발생했을 때 사건의 주체인 ‘원님’이 존재한 공간)은 ‘고향’이라고 보기 어려우므로, ‘고향’은 ‘보냈어요’에 대한 공간 표현으로 판별하지 않는다.

예) 고양이는 물고기 뱃속을 열어보았어EVENT.

- ‘뱃속’은 ‘열어보다’의 주체인 ‘고양이’가 존재하는 공간이 아닌, 단지 ‘열어보다’의 대상을 지칭할 뿐이므로, ‘뱃속’은 ‘열어보았어’에 대한 공간 표현으로 판별하지 않는다.

예) 물고기 뱃속SPACE 에 구슬이 들어있었어EVENT.

- ‘뱃속’은 ‘들어있다’의 주체인 ‘구슬’이 존재하는 공간을 지칭하므로, ‘뱃속’을 ‘들어있었어’에 대한 공간 표현으로 판별한다.

5. 사건-공간 참조의 애매성 분석을 위한 단계별 주석

서론에서 논의한 것과 같이 사건 표현과 그것이 참조하는 공간 표현은 문장 내에서 형태-통사적으로 긴밀하게 연결되어 있을 수도 있고, 그렇지 않을 수도 있다. 따라서 후자의 경우 앞뒤 문맥 또는 전체적인 맥락을 통해 주어진 사건 표현이 참조하는 가장 적절한 공간 표현을 선택해야 한다. 본 연구에서는 이에 대하여 시험 주석을 수행하였는데, 그 결과 사건 표현이 참조하는 공간 표현의 존재 여부가 명확하지 않거나, 둘 이상의 공간 표현 사이에서 하나를 선택하기에 애매한 경우가 많았으며, 이에 대해서는 주석자 간 의견도 대체로 일치되지 않는 경향을 보였다. 이러한 이유로 본 연구에서는 주어진 사건 표현에 대한 공간 표현의 판별이 어느 정도로 명확한지에 따라 세 단계로 나누어 단계별 주석을 수행하였다. 이는 주석이 완결된 이후에 주석자 간 일치도(inter-annotator agreement)를 분석하고 사건-공간 표현 간 참조의 표층적 특성을 유형화하는 데 이와 같이 단계별로 주석된 결과를 이용하는 것이 도움이 될 것으로 기대하기 때문이다.

각 사건 표현에 대한 공간 표현의 주석은 아래의 기준에 따라 총 세 단계로 나누어 이루어지며, 모든 텍스트에 대하여 이전 단계의 주석이 모두 완료된 경우에만 다음 단계의 주석을 수행하도록 하였다.

(14) 단계별 주석 기준

1 단계 통사 기반 주석

사건-공간 표현이 서로 형태-통사적으로 연결되어 있어 가장 명확하게 판별이 가능한 경우

2 단계 추론 기반 주석

형태-통사적으로 직접 연결되어 있지는 않으나, 텍스트에 드러난 정보를 이용하여 추론을 통해 애매성 없이 판별이 가능한 경우

3 단계 추측 기반 주석

텍스트에 드러난 정보를 이용하여도 애매성이 존재하여 추측을 통해서만 판별이 가능한 경우

각 단계를 예시하면 아래와 같다.

(15) 단계별 주석 예시

1 단계 동생은 방망이를 들고 집_{SPACE} 으로 들어왔어_{EVENT}.

2 단계 동생은 방망이를 들고 집_{SPACE} 으로 들어왔어. 방망이를 두드리자_{EVENT} ...

3 단계 단양 고을_{SPACE} 원님은 그 소문을 들고_{EVENT} ...

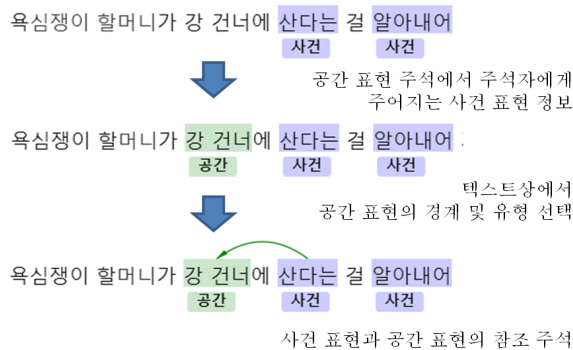
첫 번째 예문의 경우 ‘집’이 ‘들어왔어’와 통사적으로 연결되어 있으며, ‘집’이 ‘들어왔어’가 발생한 공간을 지칭하므로 1단계 주석 대상에 해당한다. 두 번째 예문의 경우 ‘집’과 ‘두드리자’가 형태-통사적으로 연결되어 있지는 않지만, 문맥상 ‘두드리자’가 발생한 공간이 ‘집’이라는 것을 애매성 없이 추론할 수 있으므로 2단계 주석 대상에 해당한다. 세 번째 예문의 경우 ‘들고’와 ‘단양 고을’이 통사적으로 연결되어 있지 않으면서, 동시에 ‘들고’가 발생한 공간이 ‘단양 고을’이라고 추정할 수 있으나 그렇지 않을 수도 있다는 해석상의 애매성이 존재하므로 (가령, 단양 고을을 관할하는 원님이 그 소문을 단양 고을이 아닌 다른 곳에서 들은 것으로 해석할 수도 있으므로) 이는 3단계 주석 대상에 해당한다. 만약 3단계의 기준으로도 공간 표현을 찾을 수 없는 사건 표현에 대해서는 공간 표현을 주석하지 않는다.

6. 주석 준비 및 주석 결과**6.1 주석 형식 및 도구**

본 연구에서는 주석 말뭉치의 구축을 위하여 형태소 단위의 주석과 격리 주석(stand-off annotation) 방식을 택하였다. 어절 단위의 주석이 주석자 입장에서 주석을 수행하기에 더 효율적인 것으로 알려져 있으나, 사건 표현의 경우 보조용언 및 접미사, 선어말어미와 같은 문법형태소들과 명사구 내 명사들을 분리하여 주석해야 하므로 형태소 단위의 주석 방식을 택하였다. 공간 표현의 경우 기본적으로 명사구를 주석하므로 문법형태소를 별개로 주석하는 경우는 없으나, 여러 개의 명사가 띄어쓰기 없이 연쇄된 경우 이를 별개로 주석해야 하기 때문에 형태소 단위의 주석 방식을 택하였다. 또한 명사구 주석 시 조사는 주석하지 않도록 하였다.

본 연구에서는 자체 제작한 웹 기반 주석 도구를 사용하여 주석자가 주석한 텍스트를 XML 형식의 파일로 출력할 수 있도록 하였다. 이 도구에서는 주석자가 텍스트 상에서 원하는 범위를 마우스로 선택한 후 그것에 특정 표지를 부착하여 주석하는 방식을 지원한다. 주석자는 텍스트의 특정 범위를 선택한 후, ‘사건’ 또는 ‘이동사건’ 중 하나의 표지를 선택함으로써 하나의 사건 표현에 대한 주석을 완료한다. 공간 표현 주석에 대해서는 사건 표현 주석과 마찬가지로 먼저 텍스트의 특정 범위를 선택하고 그것에 ‘공간’ 표지를 부착한 후, 그것을 참조하는 사건 표현을 지정하기 위해서 미리 주석된 사건 표현 중 하나를 선택하며, 이로써 하나의 공간 표현에 대한 주석을 완료한다. 만약 이동 사건 표현에 대한 공간 표현을 주석하는 경우, 공간 표현을 선택하면서 동시에 ‘출발점’, ‘경로’, ‘도착점’의 세 표지 중 하나의 참조 유형을 선택한다. 공간 표현을 주석할 때에는 항상 그것을 참조하는 하나의 사건 표현을 선택하도록 하였으며, 사건 표현이 참조하지 않는 공간 표현은 주석하지 않도록 하였다.

[그림1]은 주석 도구상에서 사건 표현과 공간 표현이 주석된 상태를 나타내고, [그림2]는 이에 대한 주석 정보를 XML 형식으로 출력한 결과를 나타낸다.



[그림 1] 사건-공간 참조 주석 과정 예시

[그림2]에서 <space> 와 <event> 는 각각 공간 표현과 사건 표현의 정보를, <relation> 은 사건-공간 표현 간 참조 정보를 담고 있다. 은 공간 표현과 사건 표현의 텍스트상의 범위 정보를 담고 있으며, <type> 은 참조 유형(공간/경로/출발점/도착점), <args> 는 해당 참조 관계에 참여하는 사건 표현과 공간 표현의 식별자 정보를 담고 있다.

6.2 데이터

본 연구에서는 사건의 흐름이 뚜렷하면서도 사건 표현에 비해서 공간 표현이 상대적으로 드물게 언급되는 동화 텍스트를 분석 대상으로 삼았으며, 아래와 같이 총 10편의 초등학생 대상 동화로 대상을 선정하여 주석을 진행하였다. 동화에서는 일반적인 도메인과 달리 의인화가 자주 일어나고 다른 장르와 달리 내용의 전개 과정이

```

<space id="129">
  <text>강 건너</text>
  <span>
    <begin>1910</begin>
    <end>1914</end>
  </span>
</space>
<event id="130">
  <text>산 닦는</text>
  <span>
    <begin>1916</begin>
    <end>1919</end>
  </span>
</event>
<event id="131">
  <text>알 아 내어</text>
  <span>
    <begin>1922</begin>
    <end>1926</end>
  </span>
</event>
<relation id="12">
  <type>space</type>
  <args>
    <event>130</event>
    <space>129</space>
  </args>
</relation>

```

[그림 2] 주석 결과 출력 예시

작가의 상상력에 크게 의존하기 때문에, ‘마을’, ‘산 속’과 같이 전형적인 공간 표현과 더불어 ‘물고기 뱃속’, ‘램프’와 같이 일반적인 도메인에서는 공간을 지칭하지 않는 표현들까지 다양한 표현들을 포함하고 있다. 주석은 원시 텍스트를 대상으로 하였으며, 형태통사적 정보와 같이 추가적인 언어학적 정보는 따로 제공하지 않았다.

- (16) 개와 고양이 / 견우와 직녀 / 도깨비와 개암 / 미운 오리 새끼 / 소가 된 게
 으름뱅이 / 선녀와 나무꾼 / 소금장수와 기름장수 / 은혜 값은 까치 / 잣과
 콩나무 / 헨젤과 그레텔

[표 1]은 주석 말뭉치의 어절 개수, 문장 개수 등의 통계를 나타낸다. 평균 462개의 어절, 65개의 문장으로 구성되어 있으며, 각 문장은 평균 7개의 어절로 구성되어 있다.

제목	어절 개수	문장 개수	내레이션 (발화문 제외) 문장 개수	문장 내 평균 어절 개수	내레이션 문장 내 평균 어절 개수
개와 고양이	787	88	57	8.6	8.5
견우와 직녀	373	48	34	7.6	8.2
도깨비와 개암	369	61	42	5.9	6.1
미운 오리 새끼	435	76	54	5.6	6.1
선녀와 나무꾼	435	65	47	6.5	6.8
소가 된 게으름뱅이	662	67	46	8.3	8.4
소금장수와 기름장수	447	63	44	6.6	6.8
은혜 깊은 가치	411	57	36	6.4	6.2
잭과 콩나무	349	65	47	5.0	5.3
헨젤과 그레텔	352	58	39	5.8	5.9
계	4620	648	446	-	-
평균	462	65	45	6.6	6.8

[표 1] 데이터 세부 정보

6.3 주석 진행 및 결과

사건 표현은 문법적으로 비교적 명확하게 드러나는 자질에 따라 주석하도록 하였으므로, 주석자 간 불일치가 거의 발생하지 않을 것으로 예상하여 1명의 주석자가 진행하고 다른 1명의 주석자가 이를 검토하도록 하였다. 공간 표현 주석의 경우 주석자 간 불일치의 가능성을 고려하여 2명의 주석자가 동시에 주석을 진행하였다. 두 주석자는 모두 전산언어학을 전공하고 한국어를 모국어로 사용하는 연구원으로서, 언어학에 대한 기본 배경 지식을 가지고 있어 본 연구의 목적과 주석 가이드라인을 이해하고 주석을 진행하는 데 문제가 없는 것으로 판단하였다. 또한 5절에서 논의한 대로 공간 표현 판별의 애매성에 따라 주석을 세 단계로 나누고, 10개의 텍스트에 대해 현재 단계 주석이 모두 완료된 이후에 다음 단계 주석으로 넘어가도록 하였다. 단, 주석자의 주석 가이드라인에 대한 숙지 정도 파악 및 훈련을 위해 각 단계의 텍스트를 1개, 3개, 6개의 세 묶음으로 나누고, 각 묶음 내 텍스트의 주석이 완료되면 두 주석자가 회의를 개최하여 주석자 간 일치도를 확인하고 불일치한 공간 표현에 대해서는 주석 가이드라인에 따라 서로 합의하여 하나의 공간 표현을 선택하게 하였다. 즉, 회의가 끝나면 주석자 간에 합의된 주석 말뭉치가 생성되며 다음 묶음의 텍스트에 대한 주석으로 넘어가게 된다. 단, 3단계 주석의 경우 1, 2단계와는 다르게 추측을 기반으로 판별이 이루어지기 때문에 주석자 간 합의가 큰 의미가 없다고 판단하여 각 주석자가 판별한 공간 표현을 모두 주석 말뭉치에 포

합하게 하였다. 이러한 방식으로 10 개의 텍스트에 대해 단계별로 세 번의 회의를 개최하여 총 9 번의 회의가 개최되었다. 주석자 간 일치도는 다음과 같이 계산한다.

(17) 주석자 간 일치도 =

$$\frac{\text{주석자 간 공간 표현 주석(공간 표현의 범위 및 참조 유형)이 일치한 사건 표현의 개수}}{\text{전체 사건 표현의 개수}}$$

[표2]는 주석 결과를 나타낸다. 단계별 주석자 간 일치도를 제외한 다른 수치들의 경우 앞 단계의 주석을 모두 포함한 수치이다. 공간 표현 개수와 사건-공간 참조 개수의 경우, 1, 2 단계에서는 불일치한 참조에 대해서 주석자 간 합의를 거쳐 생성된 주석 말뭉치상에서 계산된 것을 나타내며, 3 단계의 경우 합의를 하지 않고 각 주석자가 주석한 사건-공간 참조를 모두 포함시킨 것이다.

7. 주석 결과 분석

본 절에서는 주석 과정에서 나타난 특징과 표2의 수치 및 주석 말뭉치에 나타난 특징들에 대하여 논의한다.

7.1 1 단계 주석 결과

먼저 표 2에서 공간을 참조하는 사건 표현의 개수를 살펴보면, 전체 사건 표현 중 평균 14%에 해당하는 표현만 1 단계에서 공간 표현을 참조한다고 주석되었는데, 이는 서론에서 논의한 것과 같이 텍스트상에서 사건 표현과 공간 표현이 형태-통사적으로 긴밀하게 연결된 경우가 그만큼 드물다는 것을 보여준다. 또한 1 단계에서는 사건 표현에 대한 공간 표현의 판별이 비교적 명확하여 주석자 간 일치도가 매우 높게 나타나는 것을 확인할 수 있다. 1 단계에서 가이드라인 숙지에 대한 부족으로 인한 불일치 이외에 주석자 간 이견이 있었던 경우는 아래와 같다.

(18) 마을 사람들이 수군거렸지만EVENT 눈도 깜짝 안 하는 욕심쟁이었지.

(19) 옛날 옛적, 하늘나라를 다스리는EVENT 옥황상제에게 ...

1 단계 주석에서 주석자 1은 예문 (18)에서 ‘수군거렸지만’이 발생한 공간으로 ‘마을’을 주석하였고, 예문 (19)에서 ‘다스리는’이 발생한 공간으로 ‘하늘나라’를 주석하였으나, 주석자 2는 두 사건 표현에 대해서 모두 주석할 공간 표현이 없다고 판단하였다. 이 경우는 주석자 2의 결정을 따르기로 합의하였는데, 이는 “마을 사람들이 수군거렸다”를 통해 “마을 사람들이 마을에서 수군거렸다”라는 명제를 이끌어내는 것은 1 단계의 형태-통사적인 판별 기준을 넘어서 추론 또는 추측에 해당한다고 보았기 때문이다. 마찬가지로 1 단계의 기준만으로 “하늘나라를 다스리다”를 통해 “하늘나라에서 다스리다”라고 결론 짓는 것은 어렵다고 간주하였다.

제목	전체 사건 표현 개수	공간 표현을 참조하는 사건 표현의 개수 및 비중 (단계별 누적)			공간 표현의 개수 (단계별 누적)			사건-공간 표현 참조의 개수 (단계별 누적)			주석자 간 일치도		
		1 단계	2 단계	3 단계	1 단계	2 단계	3 단계	1 단계	2 단계	3 단계	1 단계	2 단계	3 단계
개와 고양이	206	30 (15%)	118 (57%)	180 (87%)	29	29	29	30	127	253	0.97	0.72	0.50
견우와 직녀	94	5 (5%)	13 (14%)	82 (87%)	5	9	9	5	13	84	0.97	0.83	0.83
도깨비와 개암	103	14 (14%)	42 (41%)	89 (86%)	14	16	17	14	48	117	0.98	0.72	0.48
미운 오리 새끼	97	17 (18%)	64 (66%)	80 (82%)	17	22	22	17	67	92	0.94	0.61	0.61
선녀와 나무꾼	109	18 (17%)	54 (50%)	65 (60%)	18	20	21	18	59	79	0.95	0.66	0.65
소가 된 계 으름뱅이	177	14 (8%)	73 (41%)	107 (60%)	14	16	17	14	78	115	0.98	0.90	0.81
소금장수와 기름장수	110	24 (22%)	74 (67%)	78 (71%)	21	21	21	24	76	83	0.95	0.59	0.92
은혜 값은 가치	93	11 (12%)	66 (71%)	74 (80%)	11	16	16	11	70	82	0.95	0.76	0.59
책과 콩나무	92	14 (15%)	45 (49%)	78 (85%)	14	18	23	14	49	115	0.92	0.67	0.26
헨젤과 그레텔	95	23 (24%)	49 (52%)	81 (85%)	22	24	25	23	57	103	0.94	0.49	0.39
계	1176	170 (14%)	598 (51%)	914 (78%)	165	191	200	170	644	1123	-	-	-
평균	118	17	60	91	17	19	20	17	64	112	0.96	0.71	0.62

[표 2] 주석 결과

7.2 2단계 주석 결과

2단계 주석에서는 1단계에서 주석된 것을 포함하여 전체 사건 표현 중 절반 정도가 공간 표현을 참조하는 것으로 주석되었다. 2단계 주석의 경우 주석자가 인지적 부담을 가장 크게 느꼈는데, 이는 사건 표현과 공간 표현이 통사적으로 연결되어 있지 않은 상태에서 텍스트상에 있는 정보만을 이용하여 사건-공간 표현 참조를 추론할

수 있는지 여부를 결정하기가 쉽지 않기 때문인 것으로 보인다. 다음은 2단계 주석에서 주석자 간에 사건-공간 참조 주석이 일치한 경우를 예시한다.

(20) 하루는 할아버지가 강_{SPACE}에 나가_{EVENT} 늦도록 낚시_{EVENT}를 드리우고 있었지만_{EVENT} 물고기는 한 마리도 잡히지_{EVENT} 않았어요.

“거참, 이상도 하지. 이렇게 한 마리 잡히지_{EVENT} 않다니.... 오늘은 그만 낚시대를 거뒀어_{EVENT} 쟈구면_{EVENT}.”

할아버지가 허리를 펴며_{EVENT} 막 낚시대를 거두러_{EVENT} 할_{EVENT} 때 갑자기 낚시대가 휘청했어_{EVENT}.

(21) 방_{SPACE}에 들어온_{EVENT} 도깨비들은 방망이를 들고_{EVENT} 요술을 부렸어_{EVENT}. 그러자 금은보화가 마구 쏟아지는_{EVENT} 거야.

두 주석자 모두 예문 (20)에서 등장한 모든 사건이 첫 문장에 언급된 ‘강’에서 발생하는 것으로 주석하였으며, 예문 (21)에서 등장한 모든 사건이 ‘방’에서 발생한 것으로 주석하였다. 본 연구에서는 이와 같이 2단계 주석 과정에서 주석자 간 주석이 일치한 공간 표현들을 분석한 결과, 다음과 같은 공통적인 특징들을 발견할 수 있었다. 첫째, 공간 표현이 따로 언급되지 않은 상태에서 주체가 동일한 사건이 연속되어 나타날 경우, 해당 사건들은 모두 동일한 공간 표현을 참조하는 경향을 보였다. 가령, 예문 (20)에서 ‘나가’, ‘낚시’, ‘드리우고 있었지만’ 및 ‘펴며’, ‘거두러’, ‘할’은 모두 연속되어 언급된 사건이며 동시에 주체가 ‘할아버지’로 동일하다. 둘째, 연속되어 언급된 사건들이 이야기 흐름상 짧은 시간 간격으로 발생한 경우, 해당 사건들이 동일한 공간 표현을 참조한다고 간주하는 경향을 보였다. 가령, 예문 (21)의 ‘들어온’, ‘들고’, ‘부렸어’, ‘쏟아지는’은 모두 도깨비가 ‘방’에 들어온 이후 짧은 시간 안에 연속하여 발생한 사건으로 볼 수 있으므로, 두 주석자 모두 해당 사건들이 ‘방’에서 발생하는 것으로 간주하는 데 이견이 없었던 것으로 보인다. 특히 ‘쏟아지는’의 경우 짧은 시간 간격을 암시하는 시간부사 ‘그러자’의 사용으로 인해 바로 이전에 언급된 사건 ‘부렸어’와 주체(‘금은보화’)가 다름에도 불구하고 시간적으로 연속되어 발생한 사건으로 간주하였다.

마지막으로, 하나의 공간에서 발생하는 큰 범주의 사건이 언급되고, 그것을 구성하는 작은 사건들이 연속되어 언급될 때, 해당 사건들은 모두 하나의 공간에서 발생하는 것으로 간주하는 경향을 보였다. 가령, 예문 (20)에 언급된 사건들은 거의 대부분 ‘낚시’라는 하나의 큰 사건을 구성하는 일련의 행위 및 세부 사건으로 볼 수 있으므로, 동일한 공간에서 발생할 가능성이 높다고 볼 수 있다.

위 조건을 만족하는 경우 대체로 두 주석자의 주석이 일치하였으나, 다음과 같이 두 공간 표현 사이에서 선택의 애매성이 존재하여 주석하지 않은 경우가 있다.

(22) “집_{SPACE}이 무너지는 소리야. 도망쳐!”

도깨비들이 공지 빠지게 집 밖_{SPACE}으로 달아났어_{EVENT}.

그중 하나는 방망이도 팽개치고_{EVENT} 말이야.

다음 날, 동생은 방망이를 들고 집으로 돌아왔어.

위 예문에 표시된 사건 표현 ‘팽개치고’의 경우 이전 문장에 언급된 ‘달아났어’와 시간적으로 연속되어 발생한 사건으로 볼 수 있으나, 주석자 1은 그것이 ‘집’에서 발생했다고 간주하여 주석한 반면, 주석자 2는 그것이 ‘집’과 ‘집 밖’ 중 어떤 공간에서 발생했는지 명확하지 않다고 판단하여 주석을 보류하였다. 이와 같이 서로 물리적 포함 관계에 있거나 인접한 관계에 있는 두 공간을 두고 일련의 사건이 발생했을 때, 그 중 일부 사건 표현에 대해서는 하나의 공간 표현을 지정하기 애매한 경우가 많았으며, 2단계에서 발생한 주석자 간 불일치 중 상당수는 이 경우에 해당하는 것으로 나타났다. 이와 같은 불일치가 발생했을 경우, 본 연구에서는 주석자 간 합의를 하는 데 있어 보수적인 입장을 취하여 ‘팽개치고’와 같은 사건 표현에 대해서는 공간 표현을 주석하지 않는 것을 원칙으로 하였다. 물론 ‘팽개치고’에 대한 공간 표현으로 ‘집’을 주석하는 것이 이야기 흐름에 비추어 크게 위배되지는 않으나, 2단계 주석의 중요한 주석 원칙 중 하나가 선택상의 애매성이 없어야 한다는 것이므로, 위와 같이 주석자가 합의하여 애매성이 있다고 판단할 경우, 2단계에서 주석 대상에 포함하지 않았다. 아래의 예문은 이와 유사한 경우를 보여준다.

(23) 말을 마친 사람은 자신이 타고 온 거북의 등에 할아버지를 태우고 물 속SPACE으로 들어갔어요.

용궁SPACE에 도착한 할아버지를 용왕님과 왕자님이 반갑게 맞이했어요.

“할아버지, 어서 오세요. 제 목숨을 살려주셔서 정말 고맙습니다.”

“내 아들을 살려 준 은혜를 갚고자 하니 마음껏 즐기시오.”

할아버지는 처음 보는EVENT 맛난 음식에 멋진 옷을 입고EVENT 용궁의 잔치를 즐거어요EVENT.

위 예문에서 마지막 문장에 나열된 사건 표현 ‘보는’, ‘입고’, ‘즐거어요’의 경우, 그것에 대한 공간 표현으로 주석자 1은 ‘용궁’에 주석한 반면, 주석자 2는 해당 사건들이 ‘용궁’ 밖에서 발생할 수도 있는 사건으로 간주하여 ‘물 속’에 주석하였다. 이 경우도 마찬가지로 보수적인 입장을 취하여 주석자 2의 의견을 따르는 것으로 합의하였다.

7.3 3단계 주석 결과

표2에 따르면 2단계 주석을 수행한 결과 전체 사건 표현 중 평균 51%가 공간 표현을 참조하는 것으로 주석되었으나, 여전히 절반에 가까운 사건 표현들이 공간 표현에 참조되지 않은 상태로 남은 것을 확인할 수 있다. 이는 사건 표현이 참조하는 공간 표현이 텍스트에 아예 나타나 있지 않거나, 또는 텍스트에 드러난 정보만으로 사건 표현이 참조하는 공간 표현을 판별하는 데 애매성이 있는 경우가 빈번하다는 것을 의미한다. 3단계 주석에서는 주어진 사건 표현에 대하여 공간 표현을 주석하는 데 있어서 두 표현 간의 참조 여부가 명백하지 않더라도, 그것이 이야기 흐름에 위배되지 않으면서 충분히 추측 가능한 범위에 포함된다고 판단될 경우 해당 공간 표현을 주석하도록 하였다. 아래의 예문은 이러한 주석의 예시를 보여준다.

(24) 마을SPACE 사람들이 수군거렸지만EVENT 눈도 깜짝 안 하는 욕심쟁이었지.

(25) 그런데 강 건너SPACE 사는 욕심쟁이 할머니가 그 소문을 들었어요EVENT.

첫 번째 예문의 경우, 2단계 주석 과정에서는 ‘수군거렸지만’에 대한 공간 표현을 주석하지 않았으나, 3단계에서는 두 주석자 모두 ‘마을’을 공간 표현으로 주석하였다. 이는 이야기 흐름상 “마을 사람들이 수군거린” 곳이 ‘마을’일 것이라고 충분히 추측할 수 있기 때문이다. 마찬가지로 두 번째 예문에서 “욕심쟁이 할머니가 그 소문을 들은” 장소가 ‘강 건너’일 것이라고 추측할 수 있기 때문에 ‘들었어요’가 참조하는 공간 표현으로 ‘강 건너’를 주석하였다.

3단계 주석은 기본적으로 추측에 기반하여 수행되기 때문에 이전 단계들에 비하여 주석자 간 의견이 불일치하는 경우가 빈번한 것으로 나타났다. 이는 표2에서 주석자 간 일치도가 1, 2단계에 비해 현저히 떨어지는 것을 통해서도 확인할 수 있다. 따라서 3단계 주석에서는 주석자 간 불일치를 해소하여 주석을 하나로 통일하는 것이 1, 2단계에 비하여 큰 실효성을 갖지 않는 것으로 보이며, 주석자들 간의 서로 다른 추측 결과를 모두 반영한다는 차원에서 주석자 간 주석이 불일치할 경우에도 이를 모두 말뭉치에 포함시키도록 하였다. 아래 예시는 3단계 주석에서 주석자 간 불일치가 발생한 경우를 보여준다.

(26) 욕심쟁이 할머니가 강 건너SPACE 에 산다는 걸 알아내어 개가 고양이를 등에 태우고 강을 건넜어요.

하지만 욕심쟁이 할머니는 구슬을 잃어버릴까 봐 대문을 꼭꼭 닫아 걸고 있었지요.

“좋은 방법이 없을까?”

들어가지 못하고 어슬렁거리던 개와 고양이는 곶간 옆SPACE 을 지나다가 ‘찍찍’하는 쥐 소리를 들었어요.

그 소리를 들은 고양이는 쏘살같이 달려갔어요.

그리고 대장 쥐의 목덜미를 누르며EVENT 소리쳤어요EVENT.

위 예문의 마지막 문장에 언급된 두 사건 ‘누르며’와 ‘소리쳤어요’에 대해서 주석자 1은 ‘강 건너’에서 발생한 것으로 추측한 반면에, 주석자 2는 ‘곶간 옆’에서 발생한 것으로 추측하였다. 따라서 3단계 주석을 통해 생성된 주석 말뭉치에는 해당 사건 표현에 대한 공간 표현으로 ‘강 건너’와 ‘곶간 옆’을 모두 주석하도록 하였다. 실제로 문맥상 ‘곶간 옆’이 지칭하는 공간은 ‘강 건너’가 지칭하는 공간 내에 포함되어 있다고 볼 수 있는데, 3단계 주석에서는 이와 같이 포함 관계에 있는 두 공간을 두고 주석자 간에 추측이 일치하지 않는 경우가 빈번한 것을 확인할 수 있었다. 가령, 예문 (23)의 ‘보는’, ‘입고’, ‘즐겼어요’에 대한 공간 표현을 결정하는 데 있어 ‘용궁’과 ‘물 속’ 사이에서 주석자 간에 의견이 엇갈린 것도 이러한 경우에 해당한다. 이 경우 예문 (26)과 마찬가지로 3단계 주석을 통해서 ‘용궁’과 ‘물 속’이 모두 최종 주석 말뭉치에 포함되게 된다.

7.4 공간 표현을 참조하지 않는 사건 표현

3단계 주석을 통해서도 공간 표현이 주석되지 않은 사건 표현의 경우 다음과 같은 세 가지 유형으로 분류할 수 있었다.

- (27) 공간 표현을 참조하지 않는 사건 표현의 유형
- (A) 아직 발생하지 않은 사건에 대한 예측, 의도, 소망, 능력 등의 양태가 나타난 표현
 - (B) 시간 및 날씨 변화를 지칭하는 표현
 - (C) 특정 공간에 국한되지 않는 등장인물 및 사물의 속성을 지칭하는 표현

첫 번째 유형으로 아직 발생하지 않은 사건에 대한 예측, 의도, 소망, 능력 등의 양태 의미가 나타난 표현의 경우, 그것이 특정 공간 표현을 참조한다고 보기 어려운 경우가 많았다. 아래의 예시는 이를 보여준다.

- (28) 하지만 며칠이 지나자 그렇게 훌륭한 잔치도 시들해지고 할머니가 보고_{EVENT} 싶어 견딜 수가 없더라고요.
- (29) “야옹이, 너 왜 대답을 안 하는 거야? 혹시 그 구슬을 혼자서 차지_{하려는EVENT} 속셈 아니야?”
- (30) “너 나처럼 쥐를 잡을_{EVENT} 수 있니?”

예문 (28)의 ‘보고’의 경우, 두 주석자 모두 공간 표현을 지정하지 않았는데, 이는 사건의 주체가 ‘보다’라는 사건에 대한 소망을 드러내는 과정에서 특정 공간을 상정했다고 보기 어렵기 때문이다. 마찬가지로 예문 (29)에서 의도의 의미를 나타내는 ‘차지하려는’과 예문 (30)에서 능력의 의미를 나타내는 ‘잡을’에 대해서도 유사한 이유로 두 주석자 모두 공간 표현을 지정하지 않았다.

두 번째 유형인 시간의 변화를 지칭하는 표현의 경우, 시간 변화에 대한 구체적인 공간을 상정하기 어렵다고 판단하여 따로 공간 표현을 지정하지 않았다. 아래 예문에 표시된 사건 표현들이 이러한 경우에 해당한다.

- (31) 욕심쟁이 형은 그 길로 개암을 따서 낡은 집을 찾아갔어.
도깨비 방망이가 탐났던 거지.
밤이 되자_{EVENT}, 정말로 도깨비들이 나타났네.
- (32) 그런데 선비는 산속에서 길을 잃었어.
길을 찾다 보니, 금세 날이 어두워졌지_{EVENT}.
- (33) 옥황상제에게 꾸중을 들은 뒤에, 견우와 직녀는 열심히 일을 했습니다.
하지만 시간이 흐르자_{EVENT}, 다시 게으름을 피웠지요.

그 외 사건 표현이 특정 공간을 상징하지 않아 주석자가 공간 표현을 지정하지 않은 경우로 아래의 예시에 표시된 사건 표현을 들 수 있다. 주로 특정 공간에 국한되지 않는 등장인물이나 사물의 속성을 설명하는 표현이 이에 해당한다.

- (34) 나무꾼은 홀어머니를 정성껏 모시는EVENT 효자였지요.
- (35) 옛날 사람들은 칠월 칠석에 비가 내리면, 견우와 직녀가 헤어지는 것이 아쉬워서 눈물을 흘리는 거라고 여겼습니다. 또, 그 무렵에 까치와 까마귀의 머리털이 빠지는데EVENT 오작교를 놓다가 견우와 직녀의 발에 밟혀서 털이 빠진다고 생각했답니다.
- (36) “그럼 구렁이는 뭘 먹고EVENT 살아? 구렁이는 다른 짐승을 잡아 먹어야EVENT 사는EVENT 거야.”
- (37) 호랑이한테 물려 가도 정신만 바짝 차리면 산다는 말 들어 봤어EVENT?
- (38) “할멈, 이 구슬 좀 보구려. 원하는EVENT 건 뭐든 들어준대요EVENT.”

7.5 사건의 발생 여부에 따른 공간 표현 참조

각 단계의 주석 결과를 분석한 결과, 사건 표현 중 아직 발생하지 않았거나 특정 시점에 묶여있지 않은 행위 및 상태변화를 지칭하는 사건 표현들은 그렇지 않은 사건 표현들에 비하여 공간 표현을 참조하지 않은 경우가 많은 것으로 드러났다. 본 연구에서는 동화 텍스트의 이야기 흐름 내에서 실제 발생한 사건을 지칭하는 표현을 ‘발생 사건 표현’, 그렇지 않은 표현을 ‘미발생 사건 표현’으로 명명하고, 주석 말뭉치에 포함되어 있는 모든 사건 표현을 발생 사건 표현과 미발생 사건 표현 중 하나로 분류하여 그것의 공간 표현 참조 여부를 주석 단계별로 살펴 보았다. [표 3]은 이에 대한 수치를 보여준다.

발생 사건 표현의 경우 각 단계별로 전체 표현 중 평균 16%, 57%, 85%의 표현이 공간 표현을 참조하고 있는 반면, 미발생 사건 표현의 경우 평균 10%, 33%, 58%의 표현의 공간 표현을 참조하고 있는 것으로 나타나 미발생 사건 표현에 대한 공간 표현 참조가 상대적으로 더 적은 것으로 확인되었다. 또한 표에서 제시하지는 않았으나, 미발생 사건 표현의 분포를 조사한 결과 평균적으로 69%의 표현이 등장인물의 발화문 내에서 사용되는 것으로 확인되었다. 이는 발화문의 특성상 아직 발생하지 않은 사건에 대한 기대, 소망, 의도 등의 양태 의미가 자주 드러나기 때문인 것으로 보인다. 따라서 주어진 사건 표현의 공간 표현 참조 여부를 결정하는 데 있어서 해당 사건 표현이 미발생 사건을 지칭하는지 여부와 등장인물의 발화문에 사용되었는지 여부가 중요한 역할을 한다고 볼 수 있다.

8. 토의 및 결론

본 연구에서는 텍스트에 언급된 사건에 대하여 그것이 지칭하는 공간에 대한 정보가 명시적으로 드러나지 않는 경우가 빈번하다는 점에 주목하여, 주어진 사건 표현에

제목	전체 사건 표현 개수	발생 사건 표현 개수 및 비중	미발생 사건 표현 개수 및 비중	발생 사건 표현 중 공간 표현을 참조하는 표현의 개수 및 비중 (단계별 누적)			미발생 사건 표현 중 공간 표현을 참조하는 표현의 개수 및 비중 (단계별 누적)		
				1 단계	2 단계	3 단계	1 단계	2 단계	3 단계
개와 고양이	206	161 (78%)	45 (22%)	27 (17%)	98 (61%)	148 (92%)	3 (7%)	20 (44%)	32 (71%)
견우와 직녀	94	72 (77%)	22 (23%)	3 (4%)	8 (11%)	64 (89%)	2 (9%)	5 (23%)	18 (82%)
도깨비와 개암	103	77 (75%)	26 (25%)	13 (17%)	37 (48%)	66 (86%)	1 (4%)	5 (19%)	23 (88%)
미운 오리 새끼	97	85 (88%)	12 (12%)	14 (16%)	58 (68%)	73 (86%)	3 (25%)	6 (50%)	7 (58%)
선녀와 나무꾼	109	74 (68%)	35 (32%)	13 (18%)	37 (50%)	46 (62%)	5 (14%)	17 (49%)	19 (54%)
소가 된 계 으름장이	177	111 (63%)	66 (37%)	12 (11%)	65 (59%)	90 (81%)	2 (3%)	8 (12%)	17 (26%)
소금장수와 기름장수	110	82 (75%)	28 (25%)	21 (26%)	66 (80%)	70 (85%)	3 (11%)	8 (29%)	8 (29%)
은혜 값은 가치	93	64 (69%)	29 (31%)	9 (14%)	54 (84%)	58 (91%)	2 (7%)	12 (41%)	16 (55%)
쟁과 콩나무	92	73 (79%)	19 (21%)	13 (18%)	38 (52%)	65 (89%)	1 (5%)	7 (37%)	13 (68%)
헨젤과 그레텔	95	63 (66%)	32 (34%)	14 (22%)	32 (51%)	53 (84%)	9 (28%)	17 (53%)	28 (88%)
계	1176	862	314	139	493	733	31	105	181
평균	118	86 (73%)	32 (27%)	14 (16%)	49 (57%)	73 (85%)	3 (10%)	11 (33%)	18 (58%)

[표 3] 발생 사건 표현과 미발생 사건 표현의 주석 단계별 공간 표현 참조 여부

대한 공간 표현을 선택하는 과정에서 발생하는 애매성을 분석하고, 공간 표현을 선택하는 기준에 대해 논의하였다. 1절에서 제기한 세 가지 질문에 대한 답은 다음과 같이 정리된다. 첫째, 주석자 간 일치도 분석 및 주석 결과 분석을 통해 사건 표현에 대한 공간 표현을 선택하는 과정에서 애매성이 충분히 드러난다는 것을 보였다. 둘째, 본 연구에서는 주석자가 공간 표현을 선택하는 과정에서 사용하는 방법이 크게 세 가지인 것에 착안하여, 주석 과정을 3단계로 분리하여 공간 표현 선택의 애매성과 주석자 간 불일치를 각 단계별로 분석하였다. 셋째, 본 연구에서 분석한

사건-공간 표현 간 중요한 표층적 특성 중 하나로, 사건-공간 표현 간 참조가 통사적으로 긴밀하게 되어 있는 경우는 전체 사건-공간 표현 간 참조 중 14% 뿐이고, 그 외에는 모두 통사적으로 직접적으로 연결되어 있지 않다는 것이 확인되었다. 또한 전체 사건 표현 중 78%의 표현이 공간 표현을 참조하며, 나머지 사건 표현은 공간 표현을 참조하지 않는 것으로 확인되었으며, 이러한 사건 표현들은 세 가지 유형으로 분류될 수 있음을 보였다. 이외의 추가적인 표층적 특성은 본 연구에서 주석된 말뭉치를 이용하여 사건-공간 자동 식별 시스템을 구축하고 추가 실험을 수행하여 자세히 분석할 수 있을 것으로 예상된다.

본 연구에서 제작한 단계별 주석 말뭉치는 추후에 자동화된 시스템을 개발하는데 있어 학습 데이터의 구축 및 학습 자질의 선택에 유용할 것으로 기대한다. 또한 단계별 주석 말뭉치는 자연언어처리 응용 분야에서 사용자에게 필요한 정보를 맞춤형으로 제공하는 데 유용하게 활용될 것으로 예상된다. 가령, 정보 추출 및 질의 응답과 같이 비교적 명확한 정보를 제공하는 것을 목표로 하는 응용 분야에서는 1단계 또는 2단계 주석 기준에 해당하는 정보를 선별적으로 활용할 수 있을 것이며, 텍스트-삽화 자동 변환과 같이 가능한 많은 정보를 필요로 하는 응용 분야에서는 3단계 주석 기준에 해당하는 정보까지 포함하여 보다 광범위한 정보를 제공할 수 있을 것이다.

향후 연구로 좀 더 많은 데이터에 대한 주석을 진행하여 본 연구에서 제시한 사건-공간 표현 간 다양한 특징들의 통계적 유의성을 밝히는 것과, 이를 기반으로 텍스트 내 주어진 사건 표현에 대한 공간 표현을 자동으로 추출하는 연구를 계획하고 있다. 또한 본 연구에서 제시한 방법은 본질적으로 도메인에 종속되지 않은 방법이므로, 이를 뉴스 데이터에 적용하여 사건-공간 참조 형태가 동화 데이터와 어떻게 다른지를 분석하는 것도 흥미로운 주제가 될 수 있을 것으로 기대한다.

References

- 김평, 맹성현. 2006. 사건 탐지 및 추적을 위해 신문기사에서 자동 추출된 시간 정보의 유용성 판단. *정보과학회논문지: 소프트웨어 및 응용* 33(6), 564-573.
- 김평, 성기윤, 맹성현. 2003. 사건 탐지/추적을 위한 시간 정보 추출. *제15회 한글 및 한국어 정보처리 학술대회 논문집*, 22-29, 한국정보과학회 언어공학연구회.
- 유현조, 김문형, 준호 줄리아노, 남승호, 신호필. 2009. Saken: 한국어 사건 인식 시스템. *제21회 한글 및 한국어 정보처리 학술대회 논문집*, 25-30, 한국정보과학회 언어공학연구회 & 한국인지과학회.
- 유현조, 장하연, 조유미, 김윤신, 남승호, 신호필. 2011. 한국어 TimeML-텍스트의 사건 및 시간 정보 연구. *언어와 정보* 15(1), 31-62.
- 임서현, 김윤신, 조유미, 장하연, 고민수, 남승호, 신호필. 2009. KTARSQI: 한국어 텍스트의 시간 및 사건 표현 주석. *제21회 한글 및 한국어 정보처리 학술대*

- 회 논문집, 130-135, 한국정보과학회 언어공학연구회 & 한국인지과학회.
- 정영미, 김용광. 2008. 사건중심 뉴스기사 자동요약을 위한 사건탐지 기법에 관한 연구. *정보관리학회지* 25(4), 227-243.
- Chambers, Nathanael, Shan Wang, and Dan Jurafsky. 2007. Classifying temporal relations between events. In Sophia Ananiadou (ed.), *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, 173-176.
- Kolomiyets, Oleksandr, Parisa Kordjamshidi, Steven Bethard, and Marie-Francine Moens. 2013. SemEval-2013 task 3: Spatial role labeling. In Suresh Manandhar and Deniz Yuret (eds.), *Proceedings of the 7th International Workshop on Semantic Evaluation*, 255-266.
- Kordjamshidi, Parisa, Steven Bethard, and Marie-Francine Moens. 2012. SemEval-2012 Task 3: Spatial role labeling. In Suresh Manandhar and Deniz Yuret (eds.), *Proceedings of the 6th International Workshop on Semantic Evaluation*, 365-373.
- Kordjamshidi, Parisa, Martijn van Otterlo, and Marie-Francine Moens. 2010. Spatial role labeling: Task definition and annotation scheme. presented at the 7th International Conference on Language Resources and Evaluation (LREC).
- Lapata, Mirella and Alex Lascarides. 2006. Learning sentence-internal temporal relations. *Journal of Artificial Intelligence Research (JAIR)*, 27, 85-117.
- Mani, Inderjeet, Janet Hitzeman, Justin Richer, Dave Harris, Rob Quimby, and Ben Wellner. 2008. SpatialML: Annotation scheme, corpora, and tools. presented at the 6th International Conference on Language Resources and Evaluation (LREC).
- Mani, Inderjeet, Marc Verhagen, Ben Wellner, Chong Min Lee, and James Pustejovsky. 2006. Machine learning of temporal relations. In Claire Cardie and Pierre Isabelle (eds.), *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, 753-760.
- Mirza, Paramita and Sara Tonelli. 2014. An analysis of causality between events and its relation to temporal information. In Jan Hajic and Junichi Tsujii (eds.), *Proceedings of the 25th International Conference on Computational Linguistics (COLING)*, 2097-2106.
- Pustejovsky, James, José Castaño, Robert Ingria, Roser Saurí, Robert Gaizauskas, Andrea Setzer, and Graham Katz. 2003a. TimeML: Robust specification of

- event and temporal expressions in text. presented at the 5th International Workshop on Computational Semantics (IWCS).
- Pustejovsky, James, Patrick Hanks, Roser Saurí, Andrew See, Robert Gaizauskas, Andrea Setzer, Dragomir Radev, Beth Sundheim, David Day, Lisa Ferro, and Marcia Lazo. 2003b. The TIMEBANK corpus. In Dawn Archer, Paul Rayson, Andrew Wilson and Tony McEnery (eds.), *Proceedings of the Corpus Linguistics 2003 Conference*, 647-656.
- Pustejovsky, James and Jessica Moszkowicz. 2008. Integrating motion predicate classes with spatial and temporal annotations. In Donia Scott and Hans Uszkoreit (eds.), *Proceedings of the 22nd International Conference on Computational Linguistics (COLING)*, 95-98.
- Pustejovsky, James, Jessica Moszkowicz, and Marc Verhagen. 2011a. ISO-Space: The annotation of spatial information in language. In Harry Bunt (ed.), *Proceedings of the Sixth Joint ISO-ACL SIGSEM Workshop on Interoperable Semantic Annotation*, 1-9.
- Pustejovsky, James, Jessica Moszkowicz, and Marc Verhagen. 2011b. Using ISO-Space for annotating spatial information. presented at the 10th International Conference on Spatial Information Theory.
- Roberts, Kirk, Travis Goodwin, and Sanda Harabagiu. 2012. Annotating spatial containment relations between events. presented at the 8th International Conference on Language Resources and Evaluation (LREC).
- Roberts, Kirk, Michael A. Skinner, and Sanda M. Harabagiu. 2013. Recognizing spatial containment relations between event mentions. presented at the 10th International Conference on Computational Semantics (IWCS).
- Verhagen, Marc, Robert Gaizauskas, Frank Schilder, Mark Hepple, Jessica Moszkowicz, and James Pustejovsky. 2009. The TempEval challenge: Identifying temporal relations in text. *Language Resources and Evaluation* 43(2), 161-179.
- Verhagen, Marc, Roser Saurí, Tommaso Caselli, and James Pustejovsky. 2010. SemEval-2010 Task 13: TempEval-2. presented at the 5th International Workshop on Semantic Evaluation (IWCS).
- UzZaman, Naushad, Hector Llorens, Leon Derczynski, Marc Verhagen, James Allen, and James Pustejovsky. 2013. SemEval-2013 Task 1: TempEval-3: Evaluating events, time expressions, and temporal relations. In Suresh Manandhar and Deniz Yuret (eds.), *Proceedings of the 7th International Workshop on Semantic Evaluation*, 1-9

Yoshikawa, Katsumasa, Sebastian Riedel, Masayuki Asahara, and Yuji Matsumoto.
2009. Jointly identifying temporal relations with markov logic. In Keh-Yih Su, Jian Su, Janyce Wiebe, and Haizhou Li (eds.), *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*, 405-413.

접수 일자: 2014년 10월 31일

수정 일자: 2014년 12월 8일

게재 결정: 2014년 12월 15일