한국심리학회지 : 법
The Korean Journal of Forensic Psychology
2023, Vol. 14, No. 1, 49-66
http://doi.org/10.53302/kjfp.2023.03.14.1.49

자연어 처리 Triple+ 추출을 이용한 진술 일관성 판별 정확도 연구*

조 은 경[1)] 문 혜 민[2)] 윤 여 훈[3)] 전 현 정[4)] 양 기 주[†]

동국대학교 경찰사법대학 대검찰청 법과학분석과 동국대학교 정보통신공학과

성폭력 사건의 수사 및 재판 단계에서 피해자 진술의 신빙성 판단이 중요해짐에 따라 진술분석의 수요가 증가하고 있다. 피해자 진술의 일관성은 진술 신빙성 판단의 주요 기준 중 하나이다. 4차 산업혁명 시대에 점차 고도화되는 자연어처리 기술은 대화 내용을 분석하는데 확장되고 있는 점에 착안하여, 이 연구는 자연어 처리 기술인 Triple+ 추출을 적용한 진술 일관성 분석의 정확도를 확인고자 하였다. 이를 위해 진술분석 교육을 이수한 평가자가 57건의 실제 피해자 진술 녹취록에 대해 진술 일관성 분석을 실시한 후 Triple+를 이용한 진술 불일치 분석 결과와 비교하였다. 평가자의 분석 결과 확인된 18쌍의 비일관적인 문장들에 대한 Triple+를 추출하고 7가지 진술 불일치 유형으로 구분하였으며 유형별 진술 불일치 판단 규칙을 설정하였다. 분석 결과, Triple+가 평균적으로 77% 정확하게 진술 불일치를 판별하는 것으로 나타났다. 세부 유형별로는, 방향, 시점, 행동 주체 유형은 100%, 내용 부정 유형은 75%, 장소 유형은 66.7%, 사건의 순서, 피동·능동 유형 판별은 50%의 정확도로 나타났다. 또한, 무작위로 선정된 32쌍의 일관적인 문장에 대한 판단에서는 93.8%의 판별 정확도를 보였다. 이러한 연구 결과는 Triple+을 이용한 자동적 진술 불일치 판별은 진술분석의 보조도구로서 효율성을 높일 수 있을 것으로 기대된다. 인공지능 진술분석에 필요한 현존하는 자연어 처리 기술의 한계와 향후 연구의 방향에 대해서도 논의하였다.

주요어 : 진술분석, 진술 일관성, 진술 불일치, 자연어 처리, Triple+

---

* 이 논문은 2022년도 대검찰청의「진술 진위 탐지를 위한 진술분석 준거 자동화 프로그램 개발연구」연구보고서를 바탕으로 작성되었음.

1) 제 1저자: 조은경, 동국대학교 경찰행정학부 교수
2) 공동저자: 문혜민, 동국대학교 일반대학원 법심리학 전공 박사수료
3) 공동저자: 윤여훈, 대검찰청 과학수사부 법과학분석과 진술분석관
4) 공동저자: 전현정, 동국대학교 정보통신공학과 학부연구원
† 교신저자: 양기주, 동국대학교 정보통신공학과 명예교수

스팸 메일 분류부터 Apple의 Siri나 삼성 빅스비와 같은 음성 어시스턴트, 언어 번역, 포털사이트의 검색 엔진 등 자연어 처리(natural language processing)는 일상생활에 매우 밀접하게 적용되고 있다. 최근 공개된 대화형 인공지능 챗봇 ChatGPT는 사용자의 질문에 대한 알맞은 답변을 텍스트 형태로 제공할 수 있어 구글 검색 엔진을 대체할 수 있다는 호평을 받고 있다(조계완, 2023). 이처럼 문장을 이해하고 만들어내는 고도화된 자연어 처리(natural language processing, 이하 NLP) 기술이 주목을 받고, 일상에 적용되는 사례들이 점차 증가하고 있다. 이는 인간의 언어를 처리하는 기술이 점차 정확해지고 있음을 보여준다.

제4차 산업혁명 시대에 과학 기술과 법심리학의 융합이 예견되었듯이(박광배, 2020) 자연어 처리 기술의 발달은 법심리학 분야에 영향을 미치고 있다. 그중에서도 진술과 관련된 영역과 융합하는 사례를 통해 체감할 수 있다. 2019년 유럽법심리학회(European Association of Psychology and Law)에서 '인공지능'과 '인지 면담'을 접목한 진술 획득용 챗봇 연구 사례가 발표된 바 있다(Minhas, Elphick, & Shaw, 2022). 그리고 직장 내 괴롭힘이나 성차별 피해자들이 익명으로 자신의 피해를 보고하고 기록할 수 있는 챗봇 '스팟(Spot)'이 상용화되어 있다(Spot, n.d.). 국내에서는 2020년부터 NICHD (National Institute of Child Health and Human Develoment) 조사면담 기법을 적용한 성폭력 피해자 조사용 챗봇 개발이 시작되었으며(조은경, 양기주, 윤여훈, 이윤정, 문혜민, 2022), 경찰청에 음성인식 기반 피해자 조서 작성 시스템이 도입되었고(유영규, 2020), 2021년에는 성폭력 피해자의 신고를 지원하는 챗봇이 개발되었다(맹욱재, 이준환, 2021; 한상희, 신윤하, 2021). 점차 고도화되는 자연어 처리는 단순히 진술을 획득하는 것을 넘어 진술 내용을 분석하는 데에도 확대되고 있다. 최근에 진술분석의 보조 도구로서 진술분석 자동화 프로그램을 개발하여 진술분석의 자동화가 이루어지고 있다(조은경 외, 2022).

진술분석은 1950년대 독일의 형사재판에서 전문 감정을 하던 심리학자들에 의해 개발되어 객관적 증거가 부족한 아동 대상 성폭력 사건에서 피해자의 진술 신빙성을 판단하는데 중요한 기여를 하고 있다(조은경, 2010). 국내에서도 진술분석의 수요가 증가하고 있다. 전국 해바라기센터에서 활동하는 진술분석 전문가에 의한 아동 및 장애인 피해자 진술분석 건수는 연간 3천 건 이상을 상회하며, 대검찰청 법과학분석과에서도 매년 3백 건 이상의 진술분석 결과통보서가 작성되고 있다(김한균 외, 2018; 송승주, 김민지, 2021). 뿐만 아니라 각급 법원에서도 전문심리위원을 지정하여 성폭력 피해자의 진술 신빙성 평가를 의뢰하고 있다. 국내에서 활용되고 있는 대표적인 진술 신빙성 평가 방법은 독일에서 개발된 진술타당도평가(Statement Validity Assessment)인데 아동으로부터 완전하고 자발적인 진술을 수집한 뒤, 진술에 대해 준거기반내용분석(Criteria-based Content Analysis, 이하 CBCA) 적용하기, 진술에 영향을 줄 수 있는 외적 요인이나 다른 증거를 검토하기 등으로 진행된다(조은경, 2010; Vrij, 2008).

진술분석의 수요가 증가함에 따라 진술분석 전문가들의 업무 부담이 가중되어 진술분석의 효율성 혹은 신뢰도가 저해된다는 우려가 제기되기도 한다. 진술분석은 훈련받은 전문가가 진술을 수작업으로 분류하여 상당한 시간이 소요될 뿐 아니라 진술분석에 전문가의 주

관이 개입될 수 있어 분석 결과의 신뢰도가 저해될 여지가 있기 때문이다(이미선, 2018). 진술타당도평가 절차에 포함되는 CBCA 분석은 진술분석 전문가의 주관이 개입될 여지가 있는 진술 내용 분석 단계로서 준거 출현 여부에 대해 전문가마다 다른 의견을 제시하는 경우 진술분석 결과의 신뢰도가 낮아질 수 있다. 최근에 빠른 속도로 발전하고 있는 인공지능 자연어처리 기술을 CBCA와 같은 진술내용 분석에 활용한다면 현존하는 진술분석 실무의 한계점을 개선하고 진술분석에 대한 신뢰도를 높일 수 있을 것이다.

진술분석에 과학 기술을 적용하려는 흐름에 따라 이 연구에서는 진술분석에 자연어 처리 기술을 활용하는 과정에서 자연어 처리 기술의 정확도를 실증적으로 확인하고자 하였다. 진술분석은 진술 일관성, 구체성, 합리성 등에 대한 심층적인 분석을 포괄한다. 이 중에서도 진술 일관성은 진술 신빙성 판단에 핵심적으로 고려되는 기준이다(이선미, 박용철, 2020). 따라서 이 연구에서는 자연어 처리를 통한 진술 일관성 판단의 정확도를 살펴보고 향후 보다 객관적이고 자동화된 진술분석으로 발전 가능성을 모색해보고자 하였다.

진술 일관성

진술 신빙성 인정을 위해 진술 내용의 합리성, 상당성, 일관성, 구체성 등이 요구되는데, 진술 일관성은 피해자 진술 신빙성 판단의 주요 기준 중 하나이다(이선미, 박용철, 2020). 피해자 진술 신빙성과 관련된 판례분석 연구 결과를 살펴보면, 비일관적인 진술 특징이 나타나는 경우 피해자의 진술 신빙성을 인정하지 않는 경향을 확인할 수 있다(강은영, 강민영, 박지선, 2020; 공정식, 류경희, 2017; 김민희, 이승진, 2019; 이미선, 2020; 이선미, 박용철, 2020). 법원이 정의한 진술 일관성은 진술의 주요 부분이 번복되지 않고 유지되는 것을 의미한다(이선미, 박용철, 2020). 구체적으로, 최초 폭로 단계에서부터 재판 단계까지 쟁점 사실에 관한 진술의 핵심 내용에 변함이 없는 것을 뜻한다. 이때, 핵심 사실 이외의 부가적인 내용에 대한 사소한 비일관성이 진술 신빙성을 저해하지는 않는다. 즉, 핵심 사실이 변하지 않고 의미적으로 일관성을 유지할 때 진술 일관성이 있다고 할 수 있다(이형근, 김성희, 2022).

진술 일관성은 진술의 정확도나 합리성과 혼동될 수 있으나 명확히 구분되는 개념이다. 진술 일관성은 진술의 전후가 일치하는지 평가하는 반면에 정확도는 진술이 실제 사실에 얼마나 부합되는지와 관련되어 있다(이선미, 박용철, 2020). 그리고 진술의 합리성은 진술이 현실적으로 가능한지 평가하며 법원에서는 진술의 논리칙 또는 경험칙이라고 명명하고 있다(이선미, 박용철, 2020; 이형근, 김성희, 2022). 즉, 진술의 일관성은 진술 내용에 초점화되어 있다고 할 수 있다. 진술 일관성은 여러 진술 간 비교로 평가할 수 있다. 하나의 면담에서 초기에 언급한 내용과 후반부의 내용을 비교할 수 있으며 이를 진술 내 일관성이라 한다. 또한, 면담이 2회 이상(경찰 조사, 검찰 조사, 증인 신문 등) 진행되었다면 진술 간의 일관성을 확인할 수 있다. 이 외에도 다른 참고인 진술과의 일치 여부나 물리적 증거와의 진술 일치 여부를 평가 할 수 있지만, 일반적으로 진술 일관성을 평가할 때에는 진술 내 일관성 또는 진술 간 일관성을 의미한다.

진술 일관성은 그 정도를 평가하므로 진술 간 불일치가 진술의 신빙성을 배척할 수 있는 수준인지에 판단이 이루어져야 한다(이형근, 김성희, 2022). 진술 일관성의 수준은 반복, 생략, 모순 그리고 보충으로 유형화하여 살펴볼 수 있다(Vredeveldt, van Koppen, & Granhag, 2014). 반복은 말 그대로 이전 진술과 일치하는 내용을 의미하고, 생략은 이전 진술에서는 언급하였으나 이후에는 언급되지 않은 내용을 뜻한다. 모순은 이전 진술과는 상응하지 않는 내용을 말하며, 보충은 이전에는 언급하지 않았지만 이후에는 진술한 내용을 의미한다. 따라서 진술 일관성을 판단할 때 진술 간 얼마나 일치하는지, 얼마나 생략 또는 보충되었는지, 그리고 얼마나 모순되는지 등이 모두 분석되어야 한다.

진술분석에 사용되는 내용분석 방법에서도 진술의 일관성은 중요하게 고려된다. 가장 대표적으로 사용되는 CBCA에서 진술의 일관성은 '논리적 일관성' 준거로 평가된다(Vrij, 2008). '논리적 일관성'은 사건에 대한 전체적인 진술이 이전의 진술이나 객관적 사실과 배치되지 않고 논리적으로 모순되는 점이 없는지를 의미한다(문혜민, 조은경, 윤여훈, 2022). 또 다른 진술분석 방법인 RM(Reality Monitoring)에서는 '이야기의 재구성' 준거에서 논리적 일관성을 평가한다(Vrij, 2008). '이야기의 재구성'은 진술 내용이 제기된 사건을 재건할 수 있는 수준을 평가하는 준거로서, 이때 진술의 일관성이 함께 고려된다. 즉, 진술 일관성이 이야기의 재구성을 해치지 않는 수준에서 평가된다.

다양한 진술 일관성 분석 방법을 종합해보면, 진술이 일관된다는 점을 반대로 모순되거나 불일치하는 부분이 존재하지 않는다는 점을 통해 확인하는 것이라 할 수 있다. 즉, 진술에 불일치하는 내용을 기반으로 진술이 일관적인지 아닌지를 판단하는 것이다. 이 연구에서는 이러한 진술 일관성 판단 과정을 자연어 처리 기술에 동일하게 적용하였을 때 진술 불일치가 식별되는 정확도를 확인하고자 하였다.

진술분석과 자연어 처리

자연어 처리(NLP)는 컴퓨터가 인간의 언어를 이해, 생성, 조작할 수 있도록 하는 인공지능(Artificial Intelligence, 이하 AI)의 하위 분야 중 하나이다. 자연어 처리에서 '자연어(natural language)'란 우리가 일상생활에서 사용하는 언어를 의미한다(Johri, Khatri, Al-Taani, Sabharwal, Suvanov, Kumar, 2021). 인공적으로 만들어진 언어를 나타내는 인공어(constructed language)와는 대치되는 개념으로 한국어, 영어, 중국어, 일본어 등 일상생활에서 사용되는 다양한 언어들이 자연어에 해당한다. 자연어는 인간의 의사소통에서 여러 가지 의미를 담고 있다. 자연어 처리란 자연어에서의 의미를 분석하여 자연어와 자연어 속에 담긴 의미를 이해하지 못하는 컴퓨터가 자연어를 처리할 수 있도록 하는 것이다. 자연어 처리에는 정보 검색, 정보 추출, 단어 분류, 품사 태깅 등의 자연어 처리를 위한 다양한 기술들이 존재한다.

국내에서 진술분석에 자연어 처리의 적용은 언어분석 프로그램인 K-LIWC(Korean Linguistic Inquiry and Word Count)로 시도된 바 있다(김영일, 김영준, 김경일, 2016; 문옥영, 김시업, 전우병, 김범준, 2011; 유연재, 2016). K-LIWC는 분석 대상 문서에서 형태소, 일반 명사, 조사의 언어적 지표와 감정·정서적 과정, 긍정

적인 정서, 긍정적인 느낌, 감각・지각적인 과정을 분석하는 프로그램이다. 하지만 K-LIWC는 진술 신빙성 판단의 주요 기준인 일관성, 구체성, 합리성 등을 추출하도록 설계되지 않았기 때문에 진술 신빙성 판단에 즉각 활용될 수 있다고 보기는 어렵다. 그럼에도 K-LIWC를 활용한 꾸준한 연구들은 진술분석에 자연어 처리 기술 활용 수요가 높아지고 있음을 보여준다.

최근에 인공지능 자연어 처리 기술과 CBCA 준거 분석의 융합 연구가 시도되기도 하였다(신준호, 신정수, 조은경, 윤여훈, 정재희, 2021; 신준호, 신정수, 조은경, 윤여훈, 정재희, 2022). 신준호 외(2021, 2022)의 연구에서는 익명화된 성폭력 피해자 진술 녹취록을 CBCA 준거들의 조작적 정의에 따라 인공지능 자연어 처리 모델인 RobertA와 Bert를 이용하여 분류하였다. 그 연구에서는 세부적인 진술 내용을 분석하는 일부 준거[1]에 대해서만 모델링이 이루어졌으며, 맥락상 깊이(준거 4), 상호작용 묘사(준거 5), 대화의 재현(준거 6), 주관적 심리상태 묘사(준거 12), 기억 부족의 시인(준거 15)에서만 0.8 이상의 정확도가 도출되었다. 또한 그 연구에서는 진술 신빙성 판단에 중요하게 고려되는 진술의 일관성을 판단하는 모델은 구현되지 않았다는 한계점이 있다. RobertA와 Bert 등 자연어 처리 모델은 빅데이터 기반 텍스트 마이닝을 추구하므로 소량의 데이터를 대상으로 진술분석 준거를 추출할 경우 정확도가 낮아지는 문제가 있으며 진술의 일관성을 판별하기 위해서 매우 많은 양의 데이터를 필요로 한다(Liu et al, 2019). 그러나 성폭력 피해자 진술분석 작업에서 인공지능이 학습해야 할 데이터는 일반적인 언어학습 모델 데이터와 성격이 다를 뿐 아니라 학습용 데이터를 구하기도 어려운 문제가 있다. 또한, 해당 모델은 블랙박스(black box)로 알고리즘의 결론에 대한 이유 있는 근거가 부족하여 분석 결과의 신뢰도가 떨어질 수 있다. 본 논문에서는 소량의 전문 지식 기반 라벨링(labeling) 데이터를 중심으로 하는 자연어 처리 기법인 Triple 추출을 통해 해결하여 보고자 하였다.

Triple 추출이란, 자연어 처리에서 가장 일반적인 정보 추출의 형태이다(Rusu, Dali, Fortuna, Grobelnik, & Mladenic, 2007). Triple은 문장의 구성요소 성분 중 주어, 목적어, 서술어 세 가지를 의미한다. Triple을 추출함으로써 문장에서 중요한 개체들과 개체들 사이의 관계 파악이 가능하다. 자연어 처리 기술 중 하나인 Triple 추출을 이용하여 자연어의 의미를 분석할 수 있는 것이다. 기존의 Triple은 주어, 목적어, 서술어의 3가지 정보만을 추출하여 문장의 의미를 분석한다. 본 논문에서는 기존의 문장 분석에 사용되던 Triple과 함께 추가적으로 다른 개체들을 함께 추출하여 문장의 의미를 더욱 구체적으로 분석하고자 하였다. 따라서 문장 내의 방향, 장소, 시점, 관형어, 부사어의 다섯 가지 개체를 함께 추출하기로 하였고, 이렇게 8가지 요소를 추출하는 방식을 Triple+라고 명명하였다.

방향은 오른쪽, 왼쪽, 위, 아래 등 방향을 나타내는 명사를 의미한다. 장소는 공원, 놀이터, 학교 등 장소를 의미하는 명사를 말하며, 시점은 3시, 10시, 아침, 저녁 등 시점을 나타내는 명사와 함께 밥을 먹고 있을 때, 집에 들어온 이후 등 사건이 발생한 시간을 표현하는 다양한 표현들을 포괄한다. 관형어와 부사어는 각각 문장 내 체언을 수식하는 문장 성

1) 19개 준거 중 준거 4번부터 19번까지에만 적용되었음.

분, 용언을 꾸며주는 문장 성분으로 문장 성분의 종류인 관형어, 부사어와 의미가 일치한다. Triple+는 소량의 라벨링 데이터로 모델링이 가능하고, Triple+의 판단이 블랙박스가 아닌 전문 지식에 근거하기 때문에 분석 결과에 대한 신뢰성을 보장할 수 있다. 이러한 장점과 더불어 Triple+는 더욱 구체적인 성분을 추출할 수 있으므로, 이 연구에서는 Triple+를 활용하여 다차원적으로 진술 불일치를 판별하고자 하였다.

방 법

연구 대상

이 연구는 대검찰청에서 2020년부터 추진 중인 '형사사법증거 분석기법 표준화를 위한 기반구축 연구개발' 사업에서 수집된 57건의 미성년자 또는 지적장애인 대상 성폭력 피해 진술 녹취록을 분석하였다. 모든 진술은 해당 피해자와 진술분석관이 1회 면담한 내용이었다. 모든 녹취록은 진술분석관에 의해 진술의 '신빙성 있음'으로 분류된 사건이었으며, 이름, 지명 등 개인 식별이 가능한 정보는 4명의 진술분석관이 교차 검증하여 특정 불가능한 정보로 대체된 후 연구자에게 제공되었다. 죄명은 진술분석관이 구분한 내용을 토대로 설정되었다. 한 피해자가 반복 사건을 언급한 경우 개별 사건별로 구분되어 총 88개의 사건이 확인되었다. 분석에 포함된 사건의 죄명은 표 1과 같다.

표 1. 범죄 유형별 빈도(개별 사건 기준)

범죄 유형	범죄 유형 빈도(%)	진술 불일치 빈도(%)[a]
강간	38(43.2)	9(31.0)
강제추행	38(43.2)	13(44.8)
유사강간	2(2.3)	0(0.0)
준강간	3(3.4)	3(10.3)
준강제추행	6(6.8)	4(13.8)
준유사성행위	1(1.1)	0(0.0)
전체	88(100)	29(100)

a: 평가자가 판단한 진술 불일치 빈도

평가자의 진술 일관성 코딩

선행연구에서 재정립된 CBCA 매뉴얼(문혜민, 조은경, 윤여훈, 2022)[2]에 따라 훈련받은 6명의 평가자가 88건의 녹취록의 진술 내 일관성을 분석하였다. 평가자들은 녹취록 분석에 앞서 5시간에 걸쳐 진술분석 방법론과 재정립된 CBCA 정의 및 평가 방법에 대한 교육을 이수하였다. CBCA의 논리적 일관성(준거 1)에 따르면, 한 번의 진술 내에서 이전의 진술 내용이나 객관적 사실과 모순되는 세부정보나 진술인이 충분히 설명할 수 없는 모호한 세부정보가 있을 때 진술이 불일치한다고 평가한다. 이 연구에서는 해당 기준에 따라 진술 내용에 불일치하는 내용이 나타나는지를 분석하였다 평가자들은 2개의 모의 진술 사례를 통해 코딩 연습을 시행하였으며, 녹취록 코딩 교육 과정에서 불분명하거나 모호한 내용에

2) 문혜민 외(2022)는 CBCA 준거들의 명칭, 모호한 조작적 정의, 명확한 평가 기준을 보완하여 CBCA 준거를 재정립하였다. 재정립된 CBCA 매뉴얼에 따라 10시간 동안 교육받은 8명의 법심리학 대학원생들 간 신뢰도(ICC)는 .697-.949로 만족스러운 수준임이 확인되었다.

표 2. 평가자의 진술 일관성 코딩 예시

진술 내용
[제가 만약에 이렇게 누워 있으면 오른쪽.]{1} 그때 집 구조를 생각하면 창문이 있는 쪽이 있고, 부엌이 있는 쪽이 있는데, 부엌 쪽으로 머리를 두고 잤다는 말이에요. 부엌 쪽으로 머리를 두고 잤으며 [제 왼쪽에 있었어요.]{1}
아빠가 그때는 콘돔 사러 안 갔고 그냥 [바로 그때 밑에 바닥에 쌌어요, 그때처럼 똑같이. 싸고]{1} 넣고 나서 모자는 하다가, 거울 때문에 걸리적거려서 모자가 밖으로 벗겨지고 하다가 더웠는지 위에 옷 다 벗고 자기 가슴 만져달라고 그래서 손 대고 있고, 남자들은 가슴 만지면 좋아한다고 나중에 성인 돼서 하면 그래 봐라 하면서 그러고, 또 하다가 밑에 저번처럼 [허벅지 쪽에 싸고]{1}

대해서는 평가자 간 합의를 통해 이견을 조율하였다. 6명의 평가자는 3개의 녹취록을 각각 코딩하였고, 이에 대한 평가자간 신뢰도(Intra-class Correlation Coefficient; ICC(2,1))는 0.841(p<0.001)로 도출되어 높은 일치율을 보였다. 자연어 처리를 위해 각 평가자는 녹취록 전체 내용을 검토한 뒤, 진술 내 불일치한 내용이 확인되는 부분에 []를 표시하고 {1}로 코딩하였고,[3] 일치하는 진술에는 별도로 표기하지 않았다. 평가자의 진술 일관성 코딩 방식은 표 2에 제시되어 있다. 평가자가 진술 불일치를 분석한 결과 총 29쌍(58개 문장)의 진술 불일치가 확인되었으며 범죄 유형별 진술 불일치 빈도는 표 1에 제시하였다.

Triple+를 이용한 진술 불일치 분석

프로그래밍

먼저, Triple+를 이용하여 진술 불일치 여부를 확인할 수 있는 프로그램을 구축하였다. 프로그램은 1) 전체 녹취록에서 피해자 답변만을 추출하는 데이터 전처리, 2) 문장분석을 통한 Triple+ 추출, 3) 해당 문장의 진술 불일치 여부 판단, 3) 결과 출력(프로그램이 일치하다고 판단한 경우 [일치], 불일치하다고 판단한 경우 [불일치]라고 제시하며, Triple+ 결과와 함께 출력) 과정으로 진행되도록 구성하였다.

데이터 전처리

평가자가 진술이 불일치하다고 분석한 데이터는 녹취록 형태로 Triple+를 추출하는 자연어 처리 과정에 바로 적용하기에는 어려움이 있다. 따라서 Triple+ 추출 과정에서 처리가 용이하도록 데이터를 재구성하는 전처리 과정이 필요하다. 전처리는 컴퓨터 언어 python 3.8.10을 이용해 코드를 구축하여 분석관의 질문은 제외하고 피해자의 답변만을 수집하는 전처리를 실시하였다. 전처리 과정을 마친 진술 데이터의 예시는 그림 1과 같다.

전처리 코드 구축은 re모듈과 replace(), splitlines() 등의 python 내장함수 또는 메소드를 이용하여 준비된 데이터의 형태에 맞게 구축하였다.[4]

3) {1}의 의미는 CBCA 준거의 첫 번째 준거(논리적 일관성)에 해당한다는 의미로 붙여진 것이다.

4) 전처리 코드는 데이터의 형태에 따라 달라지기 때문에 각자 준비된 데이터에 맞게 준비하여 적용하여야 한다.

민우 오빠랑 언니랑 저랑 셋이서. 민우 오빠가 말, 말하는 거예요, 저한테. 계속 말로 저한테 했는데 저도 약간 인사했는데 그다음에 인사 끝나고 우리 집 구경시켜준 거예요.
친언니. 그래서 방 구경시켜주고 그다음에 언니가 민우 오빠한테 배고프냐고 뭐 해준다고 그랬는데. 그래서 민우 오빠가 라면 먹겠다고.
언니가 라면 끓여주고 밥하고. 그래서 밥이랑 라면 완성, 아니 다, 다 하고 같이 들어가서 식탁, 거기 둘이 먹고. 저는 침대에서 앉아있었어요.
제 방으로. 제 방 들어가 가지고 둘이 같이 먹고.

그림 1. 전처리를 마친 데이터 예시

문장 분석

전처리를 마친 데이터에 대해 Stanza(Qi, Zhang, Zhang, Bolton, & Manning, 2020)를 이용하여 문장 분석을 진행하였다. Stanza는 Stanford NLP Group에서 개발한 오픈 소스 자연어 처리 도구로서 66개의 언어를 지원하며, 토큰화, 표제어 추출, 품사 태깅 등 다양한 자연어 처리 기능을 제공한다. Stanza의 문장 분석 결과를 바탕으로 주어, 서술어, 목적어, 관형어, 부사어, 방향, 장소, 시점 정보를 추출하였다. 추출한 Triple+의 예시는 표 3과 같다.

그림 2는 Stanza의 문장 분석 결과 예시이다. 해당 문장 분석 결과를 바탕으로 문장의 주어, 목적어, 서술어의 Triple 및 관형어, 부사어, 방향, 시점, 장소의 추가적인 정보를 파악하여 Triple 형태로 변환한다.


```
[
  [
    {
      "id": 1,
      "text": "아저씨가",
      "lemma": "아저씨+가",
      "upos": "NOUN",
      "xpos": "NNG+JKS",
      "head": 2,
      "deprel": "nsubj",
      "start_char": 0,
      "end_char": 4
    },
    {
      "id": 2,
      "text": "올렸거든요",
      "lemma": "올리+어+거+았+든요",
      "upos": "VERB",
      "xpos": "VV+EP+EF",
      "head": 0,
      "deprel": "root",
      "start_char": 5,
      "end_char": 10
    },
    {
      "id": 3,
      "text": ".",
      "lemma": ".",
      "upos": "PUNCT",
      "xpos": "SF",
      "head": 2,
      "deprel": "punct",
      "start_char": 10,
      "end_char": 11
    }
  ]
]
```


그림 2. Stanza 분석 결과 예시

표 3. 문장 분석 예시

분석 대상	문장 1	제가 만약에 이렇게 누워 있으면 오른쪽.
	문장 2	그때 집 구조를 생각하면 창문이 있는 쪽이 있고, 부엌이 있는 쪽이 있는데, 부엌 쪽으로 머리를 두고 잤다는 말이에요. 부엌 쪽으로 머리를 두고 잤으며 제 왼쪽에 있었어요.
분석 결과	문장 1	‘주어’: [‘제’], ‘목적어’: ‘X’, ‘서술어’: [‘눕다’, ‘누워 있다’], ‘관형어’: X, ‘부사어’: ‘이렇게’, ‘방향’: [‘오른쪽’], ‘장소’: ‘X’, ‘시점’: ‘X’
	문장 2	‘주어’: [‘창문’, ‘쪽’, ‘부엌’, ‘쪽’], ‘목적어’: [‘구조를’, ‘머리를’, ‘머리를’], ‘서술어’: [‘생각하면’, ‘있다’, ‘있다’, ‘있다’, ‘두다’, ‘두다’, ‘자다’, ‘말이에요’, ‘두다’, ‘자다’, ‘있다’, ‘있다’], ‘관형어’: ‘X’, ‘부사어’: [‘그때 부엌’, ‘부엌’], ‘방향’: [‘쪽이’, ‘쪽이’, ‘부엌 쪽으로’, ‘부엌 쪽으로’, ‘왼쪽에’], ‘장소’: [‘집’, ‘부엌’, ‘부엌’, ‘부엌’], ‘시점’: ‘X’

표 4. 진술 불일치 유형

유형	정의	빈도 (%)	Triple+ 추출 적용 사례 수 (%)
방향	동일 행위에 대해 피해자가 서술하는 방향이 다를 경우	3 (10.34)	3 (16.67)
장소	동일 사건/행위에 대해 피해자가 서술하는 사건의 장소가 다른 경우	3 (10.34)	3 (16.67)
시점	동일 사건/행위에 대해 피해자가 서술하는 사건의 시점이 다른 경우	2 (6.90)	2 (11.11)
행동 주체	피해자의 진술에서 같은 행동의 주체가 다른 경우	2 (6.90)	2 (11.11)
사건의 순서	피해자의 진술에서 사건의 순서가 일치하지 않는 경우	2 (6.90)	2 (11.11)
피동・능동	피해자의 진술에서 같은 사건에 대해 피동을 능동으로, 능동을 피동으로 진술한 경우	2 (6.90)	2 (11.11)
행위 수정	피해자의 진술에서 같은 행위를 다르게 묘사한 경우	9 (31.03)	0 (0.00)
내용 부정	피해자의 진술에서 같은 사건에 대해 부사 “안”을 붙여 부정하거나 다른 내용으로 수정한 경우	6 (20.69)	4 (22.22)
	합계	29(100.00)	18(100.00)

진술 불일치 분석

평가자가 분석한 진술 일관성 코딩을 바탕으로 진술 불일치가 확인된 내용들을 유형화하였다. 진술 불일치 유형은 총 8가지로 분류되었다(표 4 참조). 그러나 현재의 Triple 추출 기술로 불일치 판단이 어려운 경우, 즉, 유의어를 판단하는 인공지능 사전 구축 및 생략된 주요 문장 구성성분 복원 등 추가적인 기술이 적용되어야 불일치 판단이 가능한 경우는 본 연구의 분석에서 제외하였다. 구체적으로, 대명사의 해석에 따라 불일치 결과가 달라지는 경우 2쌍[5], 추출된 Triple을 통해 사실을 유추하여 새로운 Triple을 만들어 비교하여야 진술 불일치를 판단할 수 있는 경우[6] 4쌍, 포함 관계(예: 전부-일부)를 파악해야 진술 불일치를 판단할 수 있는 경우[7] 2쌍, 주요한 Triple의 요소가 생략되어 있어 진술 불일치 판단이 어려운 경우 3쌍[8]을 제외하였다. 따라서 총 18쌍

5) 대명사의 해석에 따라 진술 불일치 여부가 확정되는 사례로, 대명사에 대한 정확한 라벨링이 불가능하여 데이터로 활용할 수 없기에 해당 사례들을 제외하였다.

6) 이 연구에서 고안한 자연어 처리 모델은 Triple 추출에 한정되며, 추출된 Triple을 기반으로 새로운 사실을 유추(예: 동일한 상황인지 판단)하는 모델은 구축되지 않았다. 따라서 이 연구에서 행위 수정으로 분류된 모든 사례를 제외하였다.

7) Triple+는 주어, 목적어, 서술어, 관형어, 부사어, 방향, 시점, 장소의 정보를 추출한다. 추출한 해당 정보들만으로는 포함 관계를 파악할 수 없어 포함 관계에서 불일치가 일어나는 경우를 제외하였다.

8) 문장에서 생략된 Triple를 복원하는 모델이 요구

표 5. 진술 불일치 유형별 판단 규칙

유형	판단 근거	판단 규칙
방향	서술어, 방향	두 문장의 서술어가 같으면서 방향 불일치
장소	주어, 서술어, 목적어, 장소	두 문장의 주어, 서술어 또는 목적어가 동일하면서 장소 불일치
시점	주어, 시점	두 문장의 주어가 동일하지만 시점 불일치
행동 주체	주어, 서술어	두 문장의 서술어가 동일하지만 주어 불일치
사건의 순서	접속사(선후관계) 전과 후의 정보	첫 번째 비교 문장의 선후관계 연결어미 전 정보가 두 번째 비교 문장의 선후관계 연결어미 후 정보와 동일, 첫 번째 비교 문장의 선후관계 연결어미 후 정보가 두 번째 비교 문장의 선후관계 연결어미 전 정보와 동일
피동・능동	서술어	두 문장의 같은 행위의 서술어가 피동(능동)에서 능동(피동)으로 변경
내용 부정	서술어, 부사어	두 문장의 동일 서술어에 부사어('안')의 추가 또는 생략

의 진술 불일치 사례를 자연어 처리에 활용하였다.

추출된 Triple+의 결과와 확인된 불일치 유형을 함께 이용하여 진술 불일치를 최종적으로 판단하였다. 진술 불일치 판단은 Triple+가 추출한 주어, 목적어, 서술어, 방향, 장소, 시점, 관형어, 부사어 정보를 이용하되 진술 불일치에 핵심적으로 고려된 정보는 불일치 유형별로 상이하였다. 각 유형별로 불일치 판단에 활용된 규칙은 표 5와 같다.

결 과

Triple+ 진술 불일치 판단 정확도

평가자가 진술 일관성을 분석한 내용과 Triple+ 추출로 진술 불일치를 판별한 결과를 비교한 결과, 평균적으로 77%의 판별 정확도를 보이는 것으로 확인되었다. 진술 불일치 유형별로 살펴보면, 방향, 시점, 행동 주체 유형에서는 평가자의 분석과 Triple+의 분류 결과가 완전히 일치하였다. 그러나, 내용 부정 유형에서는 평가자가 분석한 4건 중 3건만이 불일치로 판단되어 75%의 정확도를 보였으며, 장소 유형에서는 평가자가 판단한 3건 중 2건만이 불일치로 분류되어 66.6%의 정확도를 나타내었다. 사건의 순서, 피동・능동 유형에서는 2건 중 1건만이 정확하게 판별되어 50%의 정확도를 나타내었다(표 6, 전체 사례는 표 8 참조).

Triple+ 진술 일관성 판단 정확도

Triple+가 불일치하는 내용을 불일치하다고 판단하는지에 더해 일관된 내용을 일관적이라

되는 경우로, 이 연구에서는 생략어 복원 모델을 적용하지 않아 해당 사례를 제외하였다.

표 6. 진술 불일치 유형별 Triple+ 판별 정확도

유형	평가자 분석 빈도	Triple+ 진술 불일치 판단 빈도	판별 정확도 (%)
방향	3	3	100.00
장소	3	2	66.67
시점	2	2	100.00
행동 주체	2	2	100.00
사건의 순서	2	1	50.00
피동, 능동	2	1	50.00
내용 부정	4	3	75.00

표 7. 평가자의 진술 일관성 분석 결과에 대한 Triple+의 판별 정확도

		Triple+ 판단 결과	
		일치 (%)	불일치 (%)
평가자 분석 결과	일치 (%)	30 (93.75)	2 (6.25)
	불일치 (%)	4 (22.22)	14 (77.78)

표 8. 진술 불일치 유형별 Triple+ 판단 근거

불일치 유형	번호	진술 내용	판단 결과	판단 근거[a]	오류 여부
방향	1	제가 만약에 이렇게 누워 있으면 **오른쪽**. 그때 집 구조를 생각하면 창문이 있는 쪽이 있고, 부엌이 있는 쪽이 있는데, 부엌 쪽으로 머리를 두고 잤다는 말이에요. 부엌 쪽으로 머리를 두고 잤으며 제 **왼쪽**에 있었어요.	불일치	오른쪽, 왼쪽	
	2	아빠가 그때는 콘돔 사러 안 갔고 그냥 바로 그때 밑에 **바닥에** 쌌어요, 그때처럼 똑같이. 싸고 넣고 나서 모자는 하다가, 거울 때문에 걸리적 거려서 모자고 밖으로 벗겨지고 하다가 더웠는지 위에 옷 다 벗고 자기 가슴 만져달라고 그래서 손 대고 있고, 남자들은 가슴 만지면 좋아한다고 나중에 성인 돼서 하면 그래 봐라 하면서 그러고, 또 하다가 밑에 저번처럼 **허벅지 쪽에** 싸고	불일치	바닥에, 허벅지 쪽에	
	3	**위로** **안으로**	불일치	위로, 안으로	
장소	4	그 집이 **15층**이더구요. **21층**인데	불일치	15층, 21층	
	5	그 **목욕탕**이라는 곳에서 그쪽에 다 애들 내리나 봐요. 친구들이 내렸던 곳은 **태권도**예요, 태권도 근처.	일치	목욕탕, 태권도 비교 불가	○

표 8. 진술 불일치 유형별 Triple+ 판단 근거 (계속)

불일치 유형	번호	진술 내용	판단 결과	판단 근거[a]	오류 여부
장소	6	아니, 뭐라는 거야. 네, 네. 동생은... **안방에** 있었던 걸로 생각했었는데... 동생은 **거실에** 있었고	불일치	안방, 거실	
시점	7	그건 **저녁에**, 아무도 없을 때. 그거 **아침에**, 한 10시나 11시쯤 된 것 같아요.	불일치	저녁에, 아침에	
	8	술 먹고 이렇게 있는데 제가 아빠 방에 잔다 하고 이렇게 **누워 있었는데** 제가 **TV보고 있을 때**.	불일치	누워 있었는데, TV보고 있을 때	
행동 주체	9	그래서 보지 말라 그러고 올렸어요, **내가**. **아저씨가** 올렸거든요.	불일치	내가, 아저씨가	
	10	**저가** 놀래가지고 "삼촌 뭐하는 거예요? 지금 변태예요?" 그래서 **미희가** 봐가지고 "삼촌! 지금 뭐하는 거예요? 변태예요?" 미희가 이렇게 말했어요.	불일치	저가, 미희가	
사건의 순서	11	그래서 제가 이불에다 닦았거든요. 그 뒤에 전화를, 신고를 한 거예요. 한참 있다가 나중에는 일하러 나가야 된다면서 좀 이따 온다고 닭 해놓은 것 먹으라고 했는데 전 안 먹었어요.그 사람 먹다가 나갔어요. 그러고 나서 신고한 거예요.	불일치	제가 이불에 닦은 뒤에 신고함, 그 사람이 먹다가 나간 뒤에 신고함	
	12	저 잡지를 그 **빨아먹고. 그 다음에 손을 손으로 만진** 그 다음에 **손으로 건든** 다음에 그 먼저 **짬지 만지고. 그 다음에 빨아먹고**.	일치	만지다, 건들다 의미 비교 불가	○
피동, 능동	13	제 방에 이렇게 있었는데 **누워 있었거든요** 아빠가 저를 **눕히더라고요. 눕히더만**	불일치	누워 있었거든요, 눕히더만	
	14	아빠가 깨워가지고 **삼촌 등에 업혔어요**. 그냥 삼촌 걸어가고 있는데, **뛰었어요, 삼촌 등에**.	일치	업히다, 등에 뛰다 의미 비교 불가	○
내용 부정	15	손으로 이렇게. 손을 **안 사용했어요**.	불일치	안 사용했어요	
	16	댔어요. **안 대서. 안 대서. 안 대서** 그냥 이렇게 한 거예요.	불일치	안 대서	
	17	**안 닿았어요**. 혀만.	일치	서술어 파악 불가	○
	18	**엄마는 안 계셨던 것 같은데**... 엄마는 안 계셨던 것 같아요. *** 가서 친구랑 술 마시고 있었거든요, 친구들이랑. 그래서 엄마는 안 계셨고 **엄마는 안방에 계셨고요**. 엄마랑 새아빠랑 같이 자고 있다가 온 것 같아요.	불일치	계셨다	

a: 오류 사례의 판단 근거는 연구자가 임의로 해석한 사항임.

고 판단하는지 확인하기 위해 추가분석을 실시하였다. 이를 위해 녹취록에서 평가자가 불일치하다고 표시하지 않은 문장을 무작위로 32쌍(64개 문장)을 선별하였다. 선별된 문장에 대해 Triple+로 일치 여부를 확인한 결과 32쌍 중 30쌍이 정확하게 판별되었고, 2쌍은 불일치한 것으로 분류되었다. 종합적으로, 평가자의 진술 불일치/일관성 분석 결과에 대한 Triple+의 진술 불일치/일관성 판별 정확도는 표 7과 같다. 표 8은 Triple+ 추출로 진술 불일치를 판별한 결과를 정리한 표이다. 진술 내용, Triple+ 추출 모델로 판단한 판단 결과, 진술에서 판단의 근거가 된 단어, 오류 발생 여부를 나타낸다.

논 의

본 연구에서는 최근 인공지능 분야에서 두각을 나타내는 자연어 처리 기술을 진술분석에 적용할 경우 판별 정확도를 실증적으로 살펴보고자 하였다. 총 57건의 개인 정보 비식별 처리된 성폭력 피해자 진술 녹취록에 대해 CBCA 분석 훈련을 받은 평가자가 진술의 일관성을 먼저 분석하였고, 해당 결과를 Triple+ 추출로 정확하게 판별할 수 있는지 검증하였다. 분석 결과, Triple+가 약 77%의 정확도로 진술 불일치를 판단할 수 있는 것으로 확인되었다. 세부적으로, 방향, 시점, 행동 주체 유형에서 진술 불일치는 Triple+도 100% 확인해주었으며, 내용 부정 유형에서는 75%의 정확도, 장소 유형에서는 66.67%, 사건의 순서, 피동·능동 유형은 약 50%의 정확도로 진술 불일치를 판별할 수 있는 것으로 확인되었다. 또한, 진술이 일관된 문장들에 대해서 Triple+의 분류 정확도는 약 93.75%로 나타났다. 이러한 결과는 자연어 처리 기술을 적용하여 비교적 정확하게 진술 일관성을 분석할 수 있음을 보여주었다.

자연어 처리를 이용한 진술 분석은 진술분석의 효율성을 향상시킬 수 있는 방안이 될 수 있다. 진술 일관성 분석은 진술분석 전문가가 피해자 진술 내용 전체를 숙지한 뒤, 세부내용을 면밀히 분석하기 때문에 상당한 시간이 소요되는 작업이다(문혜민 외, 2022). Triple+와 같은 자연어 처리 기술을 이용할 경우 녹취록 내용을 빠르게 파악하여 진술이 불일치하는 부분을 확인할 수 있으므로 진술분석에 소요되는 시간을 단축시킬 수 있을 것이다. 이 연구에서는 문장의 구성 요소를 토대로 진술 불일치 유형을 7가지로 세분화하였다. 진술 불일치의 유형을 나눔으로써 세세한 진술 불일치를 직관적으로 이해할 수 있게 한다는 점에서 진술의 일관성 판단에 도움이 될 것으로 생각된다. 물론 분석 결과 확인된 자연어 처리 정확도(50% - 100%)는 전문가를 대체할 수는 없는 수준임은 분명하다. 그럼에도 불구하고 자연어 처리를 통해 진술의 비일관된 내용과 유형을 미리 가늠해볼 수 있다면 진술 분석을 보조하는 기술로서 활용 가치가 있을 것이다.

나아가, 자연어 처리는 진술 분석의 정확도를 확인하는 데에도 활용될 수 있을 것이다. 현재까지 진술분석의 정확도는 진술분석의 양적인 측면에서 평가되어 왔다(이미선, 2018; Anson, Golding, & Gully, 1993; Gödert, Gamer, Rill, & Vossel, 2005; Hauch, Sporer, Masip, & Blandon-Gitlin, 2017). 즉, 복수의 전문가가 진술이 불일치하는 정도를 동일한 점수로 평가하는지에 초점을 두고 진행되어 왔다. 그러나

진술분석에는 어떤 내용이 불일치하는지 평가하는 질적인 측면이 공존한다. 전문가의 진술 일관성에 대한 양적 판단은 신뢰할 수 있는 것으로 확인되었지만 평가자마다 불일치하다고 판단한 진술 내용이 동일한지는 검증되지 않았다(이미선, 2018). 이 연구에서 활용한 자연어 처리 기술은 불일치하다고 판단한 구체적인 내용을 비교할 수 있게 해준다. 추후 자연어 처리의 분석 결과와 진술분석 전문가의 분석 결과를 비교하는 과정을 통해 전문가들의 진술분석 정확도 확인과 더불어 전문가의 역량을 강화하는 용도로도 활용될 수 있을 것이다.

한편, 본 연구를 진행하면서 기술적인 한계점이 확인되었다. 분석에서 제외된 12쌍의 문장과 분석에 포함되었으나 Triple+로 판단되지 못한 문장들은 현재까지의 자연어 처리 기술력으로는 해결할 수 없는 특성을 가지고 있다. 사람은 사건의 전체적인 내용과 다양한 정보와 맥락을 고려하여 진술 불일치를 판단할 수 있는 반면, 자연어 처리는 표면적인 내용의 일치 여부에 대해서만 평가하기 때문이다.

예컨대, 사례 5 녹취록 내용에 근거하면, 피해자는 '애들'과 '친구들'을 동일인물로 설명하고, '목욕탕 있는 곳'과 '태권도 근처'가 다르다고 하였다. 사람이라면 '그 목욕탕이라는 곳에서 그쪽에 다 애들 내리나 봐요'라는 문장과 '친구들이 내렸던 곳은 태권도예요, 태권도 근처' 라는 문장을 비교할 때, '목욕탕이라는 곳', '그쪽', '애들', '친구들', '태권도 근처'의 사전적인 의미와 내포하는 의미를 모두 비교할 수 있으므로 부르는 명칭이 다르더라도 '애들'과 '친구들'이 동일 인물이라는 점 또는 '목욕탕'과 '태권도 근처'가 실제로 다른 장소임을 어렵지 않게 인식할 수 있다. 또는 사례 12의 '만지다'와 '건들다'가 동일한 추행의 의미라는 점, 사례 14의 '등에 업혔다'와 '등에 뛰었다'의 의미가 같지만 형태가 다르다는 점을 사람은 쉽게 인식할 수 있다. 하지만 현재의 자연어 처리 기술로 유의어나 함축적인 정보들은 판단할 수 없다. 따라서 이 연구에서 단어가 다르지만 동일한 뜻을 가지는 단어들의 불일치 여부가 부정확하게 판단된 것으로 보인다. 향후 추출된 Triple+에서 유의어나 의미적 연관성을 인지할 수 있도록 하는 기술이 개발된다면 진술 불일치 판단의 정확성을 더욱 높일 수 있을 것이다.

이 연구에 사용된 데이터의 특성에 기인한 한계점도 존재한다. 첫째, 연구에 포함된 사건들이 모두 대검찰청 진술분석실에서 신빙성이 있다고 인정된 사례이기 때문에 진술 불일치가 적게 나타날 수밖에 없어 진술 불일치에 대한 상세한 규칙을 설정하기 어려웠다. 향후 신빙성이 없다고 판단된 사례를 추가하여 더 많은 진술 불일치 사례를 확보할 수 있다면, 세부적인 판단 규칙을 설정하여 보다 정확한 진술 불일치 판단이 가능할 것이다. 둘째, 피해자 면담 녹취록을 원형대로 분석하는 과정에서 문장 요소가 생략된 구어체에 대한 처리가 빈번하여 Triple+ 추출의 정확도가 저해되었다. 구어체에는 문장 요소의 생략이 빈번하게 발생하고, 대체로 주어와 서술어가 생략된다. 가령, 사례 17의 녹취록에 '혀가 안 닿았다'는 진술이 '혀가 닿았다'는 진술로 변경된 점을 확인할 수 있었다. 그러나 구어체의 특성상 첫 번째 문장에서 '안 닿았어요'라고 주어가 생략되었고, 두 번째 문장에서는 '혀만'이라며 서술어가 생략되었다. 주어, 목적어, 서술어로 이루어진 Triple의 특성상 문장 핵심 요소의 생략은 추출 정확도를 감소시킬 수밖

에 없다. 향후 생략된 문장 요소들을 복원한 후 Triple+가 추출된다면 보다 정확한 진술 불일치 식별이 가능할 것으로 생각된다.

오늘날 진술분석은 아동이나 지적장애인 대상 성폭력 사건에서 피해자 진술의 신빙성을 판단하는 근거로 활용도가 높아지고 있다. 그러나 진술분석 결과는 전문가 의견으로서 연성과학인 사회과학적 증거로 분류된다(강우예, 2017). 일부 판결문에 전문가의 진술분석 결과가 인용되더라도 자연과학적 증거의 신뢰도에는 미치지 못하는 실정이다. 이 연구는 현재의 자연어 처리 기술이 진술분석에 적용되었을 때 어느 수준의 효율성과 정확도를 보여주는지 실증적으로 검증한 것에 의의가 있다. 기술 수준과 데이터 구조의 한계에도 불구하고 자연어 처리를 이용한 진술분석의 효율성과 정확도가 비교적 높게 도출된 것을 본다면, 향후 자연어 처리기술의 발전을 토대로 진술분석 결과의 신뢰도 또한 향상될 것을 기대할 수 있다. 향후 CBCA의 다른 준거들에 대해서도 이러한 연구가 축적된다면 진술분석이 객관적 기술을 토대로 하는 융합과학으로 인식될 수 있을 것이다. 물론 CBCA 준거들이 발견된다고 하더라도 그것이 피해자 진술의 진위여부를 직접 시사하는 것은 아니다(Vrij, 2008). 하지만 자연어 처리를 활용한 진술분석은 피해자 진술의 특징을 구체적이고 정확하게 나타내줄 수 있어 사람 전문가의 평가 오류를 보완할 수 있을 뿐 아니라 수사와 재판 단계에서 사실 판단을 보조하는 도구로서 활용 가치가 높다. 법심리학과 인공지능의 융합연구의 발전으로 형사사법체계 전반에 진술분석 결과에 대한 신뢰도가 제고될 수 있기를 기대한다.

참고문헌

강우예 (2017). 형사절차상 사회과학적 증거의 증명력과 증거능력에 대한 연구 - 미국의 논의에 대한 분석을 중심으로 -. **법학논총**, 34(1), 199-232.

강은영, 강민영, 박지선 (2020). 성폭력피해자 진술의 신빙성에 대한 형사사법기관 판단 및 개선방안: 성인지감수성을 중심으로. **형사정책연구원 연구총서**, 1-352.

공정식, 류경희 (2017). 아동과 지적 장애인 범죄피해 진술의 신빙성 평가기준 - 국내 법원판례를 중심으로 -. **한국범죄심리연구**, 13(3), 11-24.

곽수정, 김보경, 이재성 (2013). 한국어 의존 파싱을 이용한 트리플 관계 추출. **제25회 한글 및 한국어 정보처리 학술대회 논문집**.

김민희, 이승진 (2019). 성범죄 피해 지적장애인의 진술에 대한 법원의 신빙성 판단 - 대법원 2017. 1. 25. 선고 2016도14989 판결 및 하급심 판결을 중심으로 -. **형사법의 신동향**, 63, 272-302.

김영일, 김영준, 김경일 (2016). K-LIWC를 이용한 비압박 상황의 거짓 태도 탐지. **인지과학**, 27(2), 247-273.

김한균, 윤해성, 박윤석, 김면기, 유승진, 정교일, 이덕규, 이창훈, 권양섭, 조은경, 김민지, 이윤정, 김대원, 이원상, 이경렬, 차종진 (2018). **첨단 과학수사 정책 및 포렌식 기법 종합발전방안 연구(I)**. 경제인문사회연구회 협동연구 총서 18-63-01. 한국형사정책연구원.

맹욱재, 이준환 (2021). 성폭력 피해자 지원 챗봇 디자인: 시나리오 기반과 대화 모델을 결합한 하이브리드 모델에 관한 실험적

연구. 한국 HCI 학회 학술대회, 368-373.
문옥영, 김시업, 전우병, 김범준 (2011). 한국어 진술서에서 책임회피 시 나타나는 거짓의 언어・심리적 특징. 한국심리학회지: 사회 및성격, 25(2), 91-111.
문혜민, 조은경, 윤여훈 (2022) 성폭력 피해 진술의 신빙성 평가를 위한 준거기반 내용분석(CBCA)의 준거 재정립 연구. 피해자학연구, 30(2), 71-99.
박광배 (2020). 제4차 산업혁명 시대의 법심리학: 새로운 도전과 과제. 한국심리학회지 일반, 39(4), 481-516.
송승주, 김민지 (2021) 한국 진술분석 보고서 및 증언에 대한 질적 평가 기준. 한국심리학회지: 법, 12(2), 223-251.
신준호, 신정수, 조은경, 윤여훈, 정재희 (2021). RoBERTa 를 활용한 CBCA 준거 분류 모델 성능 비교. 한국정보과학회 학술발표논문집, 296-298.
신준호, 신정수, 조은경, 윤여훈, 정재희 (2022). CBCA 준거 분류에서의 BERT 기반 모델 성능 비교. 정보과학회논문지, 49(9), 727-734.
유연재 (2016). 한국어 언어분석 프로그램(KLIWC)을 이용한 온라인 기만 사용후기 연구. 소비자학연구, 27(1), 69-92.
유영규 (2020. 12. 14.) 성폭력 피해 경찰에서 진술하면 AI가 받아 적는다. SBS 뉴스. https://news.sbs.co.kr/news/endPage.do?news_id=N1006121430&plink=ORI&cooper=NAVER
이미선 (2018). 성폭력 피해아동 진술신빙성 판단에 있어서 평가자간 신뢰도: 진술분석 전문가 집단을 대상으로. 한국심리학회지: 사회및성격, 32(2), 67-83.
이미선 (2020). 지적장애인 성폭력 사건 특성과 법원의 판단. 한국심리학회지: 법, 11(2), 211-239.
이선미, 박용철 (2020). 성폭력 형사사건에서 피해자 진술의 신빙성과 경험칙에 관한 연구. 사법정책연구원 연구총서, 1-232.
이형근, 김성희 (2022). 진술 신빙성 평가에 관한 탐색적 연구-성인지 감수성과 인격적 요소의 경향성을 중심으로. 경찰법연구, 20(3), 153-186.
조계완 (2023. 1. 30). 아이폰 등장에 비견될 '챗GPT 돌풍'에...국내 관련주 폭등. 한겨레. https://www.hani.co.kr/arti/economy/finance/1077474.html
조은경 (2010). 성폭력 피해 아동 진술신빙성 평가의 한계와 전망. 피해자학연구, 18(2), 53-67.
조은경, 양기주, 윤여훈, 이윤정, 문혜민 (2022). 성폭력 피해자 조사면담용 인공지능 챗봇 개발과 형사정책적 함의. 형사정책, 34(1), 111-138.
한상희, 신윤하 (2021. 8. 5). AI 챗봇이 성폭력 피해자 돕는다…증거수집・조서작성까지. 뉴스1. https://www.news1.kr/articles/?4394119
Anson, D. A., Golding, S. L., & Gully, K. J. (1993) "Child sexual abuse allegations: Reliability of criteria-based content analysis", *Law and Human Behavior, 17*, 331-341.
Gödert, H. W., Gamer, M., Rill, H. G., & Vossel, G. (2005). "Statement validity assessment: Inter rater reliability of criteria-based content analysis in the mock crime paradigm", *Legal and Criminological Psychology, 10*(2), 225-245.
Hauch, V., Sporer, S. L., Masip, J., &

Blandon-Gitlin, I. (2017). “Can credibility criteria be assessed reliably? A meta-analysis of criteria-based content analysis”, *Psychological Assessment, 29*(6), 819.

Johri, P., Khatri, S. K., Al-Taani, A. T., Sabharwal, M., Suvanov, S., Kumar, A. (2021). Natural Language Processing: History, Evolution, Application, and Future Work. In: Abraham, A., Castillo, O., Virmani, D. (eds) *Proceedings of 3rd International Conference on Computing Informatics and Networks. Lecture Notes in Networks and Systems,* Singapore: Springer

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., ... & Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. arXiv preprint arXiv:1907.11692.

Minhas, R., Elphick, C., & Shaw, J. (2022). Protecting victim and witness statement: examining the effectiveness of a chatbot that uses artificial intelligence and a cognitive interview. *AI & SOCIETY,* 1-17.

Rusu, D., Dali, L., Fortuna, B., Grobelnik, M., & Mladenic, D. (2007). *Triplet extraction from sentences*. Proceedings of the 10th International Multiconference” Information Society-IS, 8-12.

Spot. (n.d). Retrieved March 29, 2023, from https://talktospot.com/index

Vredeveldt, A., van Koppen, P. J., & Granhag, P. A. (2014). The inconsistent suspect: A systematic review of different types of consistency in truth tellers and liars. *Investigative interviewing*, 183-207.

Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. John Wiley & Sons.

1 차원고접수 : 2023. 02. 13.
심사통과접수 : 2023. 03. 28.
최종원고접수 : 2023. 03. 29.

한국심리학회지 : 법
The Korean Journal of Forensic Psychology
2023, Vol. 14, No. 1, 49-66

Accuracy of statement consistency analysis using Triple+ extractions of natural language processing

Eunkyung Jo[1)] Hyemin Moon[1)] Yeohoon Yoon[2)] Hyunjung Jeon[3)] Gijoo Yang[3)]

[1)]Dongguk university
Department of Police Administration

[2)]Supreme Prosecutors' Office
Forensic Science Division

[3)]Dongguk University
Department of Information and Communication Engineering

Demand for statement analysis is increasing as the credibility of the victim's statement becomes more important in the investigation and trial of sexual offence cases. The consistency of the victim's statement is one of the main criteria for judging the credibility of a victim. In the era of 4th industrial revolution natural language processing technology is rapidly growing to analyze conversation contents. This study tried to verify the accuracy of statement consistency analysis using Triple+ extractions, a natural language processing technology. Trained evaluators conducted a statement consistency analysis on 57 actual transcripts of victim statements and compared them with the results of statement inconsistency analysis using Triple+. The Triple+ for 18 pairs of inconsistent sentences from victim statements were extracted and classified into 7 types of statement inconsistency. The rules of determining statement inconsistency for each type were established. The results showed that Triple+ correctly identified statement discrepancies 77% on the average. For subtypes of inconsistency classification accuracy varied as 100% for the direction, timing, and action, 75% for content denial, 66.7% for place, and 50% accuracy of event sequence and passive/active type were found. 93.8% accuracy was achieved in the judgment of 32 randomly selected pairs of consistent sentences. The results of this study suggest a potential for automatic statement inconsistency discrimination using Triple+ as supplementary tool for human expert statement analysis. The limitations of the existing natural language processing technology required for artificial intelligence statement analysis and the direction of future research are discussed.

Key words : statement analysis, statement consistency, statement inconsistency, natural language processing, Triple+ extraction