

# 서술 시간(narrated time) 추정의 언어 간 검증: 한국어 문학 텍스트와 LLM의 비교

이시현<sup>1</sup>, 황태검<sup>1</sup>, 박혜승<sup>1\*</sup>

<sup>1</sup>협성대학교소프트웨어공학과, 화성, 대한민국

\*Corresponding Author: hs2000park@uhs.ac.kr

## 1. 연구 목표

본 연구는 다음의 두 가지 목적을 갖는다.

- ① 서술 시간(narrated time)이 주석처리 된 500자 길이 한국어 텍스트 1280조각을 확보하는 것. (이하, 데이터셋.)
- ② LLM이 한국어 텍스트에서 인간 수준의 서술 시간 추정이 가능함을 증명하는 것.

## 2. 연구 절차

- ① 연구 진행을 위해 80개의 작품을 선정한다.
  - A. 작품 선정은 Underwood(2018)의 방식을 차용하여, 정전(canonical), 당대의 유명 작품(popular), 대조군(비소설)의 기준을 따른다.
- ② 선정된 각 작품에서 500자 길이의 텍스트 조각을 16개 샘플링 한다.
  - A. 처음 두 조각과, 마지막 두 조각은 항상 포함하며, 나머지 12조각은 무작위 추출을 진행한다.
  - B. 작품 전문이 공개된 경우, 작품을 500자 단위로 미리 분절하고 샘플링을 수행한다.
  - C. 작품 전문이 공개되지 않은 경우, 출판 도서의 페이지 당 글자수를 통계적으로 추정하고, 샘플링을 시작할 ‘페이지 번호와 행 번호의 쌍’을 산출하여 수동으로 샘플링을 수행한다.
- ③ 샘플링 된 1280개 조각에 대하여, Underwood(2018)의 방법으로 서술 시간<sup>1</sup>(narrated time), 주관적 시간<sup>2</sup>(subjective time), 생략 시간<sup>3</sup>(ellipsis time)을 기술한다. (이하, 주석 작업.)
  - A. 이는 인간 주석자가 샘플링 된 텍스트를 읽고 직접 수행한다.
  - B. 인간 주석자는 협성대학교 문예창작과 소속의 학부생 3인이다.

---

<sup>1</sup> 작품 속에서 실제로 흐른 시간의 양. 발화라면 발화에 소요된 시간, 장면이라면 장면에 소요된 시간. (“두 사람은 식사를 하고 카페에 들어갔다.”: 약 30~60분.)

<sup>2</sup> 등장인물의 의식 속에서 회상, 상상, 기억 등으로 흘러간 시간의 양.

<sup>3</sup> 작가에 의해 명시적으로 건너뛰어진 시간. (예: “3년 후...”)

- C. 인간 주석자들은, 자신의 작업물만 볼 수 있다. (서로의 작업물을 볼 수 없다.)
- ④ 샘플링 된 1280개 조각에 대하여, LLM을 사용해 주석 작업을 수행한다.
- A. LLM의 주석 작업 결과가 인간 주석자들의 작업과 유사하다면, LLM을 통한 서술 시간 추정이 유효함을 보일 수 있다.
- B. LLM의 주석 작업 결과가 인간 주석자들의 작업과 차이를 보인다면, LLM을 통한 서술 시간 추정이 유효하지 않음을 보일 수 있다.
- ⑤ (LLM의 주석 작업 결과가 유효하다면) 원본 텍스트의 단위 길이 분절과, LLM을 통한 서술 시간의 주석 작업 까지를 포함하는 파이썬 라이브러리를 개발하여 배포한다. (fic-time)

### 3. 선정 작품 목록

다음과 같이 80개 작품을 선정하였다.

docid	author	title	firstpub	gender	type	selection_rationale
001	이인직	혈의 누	1906	M	Canon	한국 신소설의 효시. 고전적 문어체와 근대적 구어체의 과도기적 텍스트.
002	안국선	금수회의록	1908	M	Canon	우화 형식을 빌린 풍자 소설. 검열을 피하기 위한 우회적 서술 전략의 대표작.
003	이광수	무정	1917	M	Canon	최초의 근대 장편소설. '자유 연애'와 '계몽' 담론의 시발점.
004	나혜석	경희	1918	F	Canon	초기 여성 작가의 페미니즘 텍스트. 여성 주체의 자각을 다룸.
005	염상섭	만세전	1924	M	Canon	식민지 현실을 관찰하는 리얼리즘의 확립. 묘사 중심 문체의 표준.
006	현진건	운수 좋은 날	1924	M	Canon	단편 미학의 완성. 하층민의 삶과 아이러니 구조.
007	김동인	감자	1925	M	Canon	자연주의 소설의 대표작. 환경 결정론과 도덕적 타락을 다룸.
008	최서해	탈출기	1925	M	Canon	신경향파(New Tendency) 문학. 간도 이주민의 절대 빈곤과 저항.
009	나도향	병어리 삼룡이	1925	M	Canon	낭만주의적 색채와 비극적 사랑. 억압된 욕망의 표출.
010	홍명희	임궫정	1928	M	Popular	식민지 시기 최대의 대하 역사 소설. 토착어와 풍속사의 보고.
011	이기영	고향	1934	M	Canon	카프(KAPF) 계열 농민 소설. 식민지 농촌의 구조적 모순 형상화.
012	강경애	인간문제	1934	F	Popular	여성 노동자의 삶을 다룬 리얼리즘 대작. 계급과 젠더의 교차성.
013	심훈	상록수	1935	M	Canon	농촌 계몽 운동(브나로드)을 다룬 당대 베스트셀러.
014	이상	날개	1936	M	Canon	한국 모더니즘의 정점. 의식의 흐름 기법과 자아 분열.
015	박태원	천변풍경	1936	M	Canon	모더니즘 기법(세대 소설). 1930년대 경성 중산층과 하층민의 생활상.
016	이효석	매밀꽃 필 무렵	1936	M	Canon	향토적 서정성과 유토피아의 문제.
017	김말봉	짚레꽃	1937	F	Popular	1930년대 신문 연재 통속 소설의 대표작. 대중적 멜로드라마의 전형.
018	채만식	태평천하	1938	M	Canon	판소리 사설조 문체를 활용한 풍자 문학.
019	김내성	마인	1939	M	Popular	한국 추리소설의 효시이자 베스트셀러. 장르 문학의 태동.
020	방인근	마도의 향불	1930s	M	Popular	탐정 소설과 멜로가 결합된 대중 소설. 도시의 퇴폐성 묘사.
021	이태준	해방 전후	1946	M	Canon	해방 직후 지식인의 이념적 혼란을 다룬 자전적 소설.
022	채만식	눈 이야기	1946	M	Canon	해방 후 토지 개혁의 모순을 풍자.
023	김구	백범일지	1947	M	Control	독립운동가 김구의 자서전. 해방 공간의 비소설 대표 텍스트.
024	이광수	나의 고백	1948	M	Control	친일 행적에 대한 변명을 담은 에세이.
025	김내성	청춘극장	1949	M	Popular	해방 공간의 청춘과 민족주의를 다룬 대하 통속 소설.
026	황순원	학	1953	M	Canon	이념 대립을 넘어서 휴머니즘. 전쟁 문학의 서정성.
027	정비석	자유부인	1954	M	Popular	전후 개방 풍조와 여성의 욕망을 다룬 논쟁적 베스트셀러.
028	김성한	암흑의 거리	1950s	M	Canon	전후 실존주의적 지식인 소설.
029	손창섭	잉여인간	1958	M	Popular	전후의 무기력증과 인간 혐오를 극한으로 밀고 나간 작품.

030	이법선	오발탄	1959	M	Canon	전후 사회의 빈곤과 방향 상실감.
031	강신재	젊은 느티나무	1960	F	Canon	세련된 감수성의 로맨스 소설.
032	최인훈	광장	1960	M	Canon	분단 문학의 고전. 남북한 체제 비판과 제3의 길 모색.
033	전광용	카피탄 리	1962	M	Canon	기회주의적 인간상을 통한 현대사 풍자.
034	박경리	김 약국의 딸들	1962	F	Canon	통영 배경의 가족사 비극. 토속적 정서.
035	이어령	흙 속에 저 바람 속에	1963	M	Control	한국 문화를 분석한 당대 베스트셀러 에세이.
036	박경리	시장과 전장	1964	F	Canon	여성의 시각에서 본 전쟁과 생존.
037	김승옥	무진기행	1964	M	Canon	감수성 혁명. 4.19 세대의 내면과 현대적 단편 미학.
038	김승옥	서울 1964년 겨울	1965	M	Canon	산업화 초기 서울의 익명성과 소외.
039	이청준	병신과 머저리	1966	M	Canon	액자 소설 구조. 전쟁의 상처와 지식인의 환부.
040	이어령	장군의 수염	1966	M	Canon	추리 형식을 빌린 지식인 소설. 콜라주 기법.
041	전혜린	그리고 아무 말도 하지 않았다	1966	F	Control	요절한 여성 지식인의 에세이. 당대 문학 소녀들의 필독서.
042	박경리	토지	1969	F	Canon	한국 문학 최대의 대하소설. 5대에 걸친 서사.
043	박완서	나목	1970	F	Canon	전쟁의 상처와 중산층 여성의 시선.
044	황석영	삼포 가는 길	1973	M	Canon	산업화 시대의 유랑 노동자 로드 무비.
045	최인호	별들의 고향	1973	M	Popular	청년 문화와 호스티스 문학 논쟁의 중심.
046	리영희	전환시대의 논리	1974	M	Control	지식인 사회에 큰 영향을 미친 사회비평서.
047	황석영	장길산	1974	M	Popular	민중 사관에 입각한 대하 역사 소설.
048	법정	무소유	1976	M	Control	물질 만능주의를 비판한 수필집 스테디셀러.
049	박완서	꼴찌에게 보내는 갈채	1977	F	Control	박완서의 대표 수필집. 일상적 소재의 사회적 의미 확장.
050	문순태	정소리	1978	M	Canon	댐 건설로 인한 수몰민의 실형 의식.
051	조세희	난장이가 쓰아올린 작은 공	1978	M	Canon	1970년대 노동 문학의 정점. 환상적 리얼리즘.
052	이문열	사람의 아들	1979	M	Canon	종교와 신의 문제를 다룬 관념 소설 베스트셀러.
053	김홍신	인간시장	1981	M	Popular	대한민국 최초의 밀리언셀러. 사회학 응징 서사.
054	조영래	전태일 평전	1983	M	Control	노동 운동의 교과서가 된 평전.
055	조정래	태백산맥	1983	M	Canon	해방 전후사를 민중적 관점에서 복원한 대하소설.
056	서정윤	홀로서기	1987	M	Control	1980년대 시집 베스트셀러 현상의 대표작.
057	양귀자	원미동 사람들	1987	F	Canon	서울 변두리 소시민들의 삶을 다룬 연작 소설.
058	이문열	우리들의 일그러진 영웅	1987	M	Canon	권력의 속성을 우화적으로 다룬 작품.
059	신영복	감옥으로부터의 사색	1988	M	Control	무기수의 옥중 서간집. 인간에 대한 깊은 성찰.
060	양귀자	나는 소망한다 내게 금지된 것을	1992	F	Popular	대중적 성공을 거둔 페미니즘 스킬러.
061	이승우	생의 이면	1992	M	Canon	관념적이고 형이상학적인 문체. 해외에서 고평가.
062	신경숙	풍금이 있던 자리	1993	F	Canon	1990년대 내면성 문학의 대표작.
063	공지영	고등어	1994	F	Popular	운동권 후일담 문학이자 대중 멜로.
064	홍세화	나는 빠리의 택시운전사	1995	M	Control	한국 사회에 '플레랑스' 개념을 소개한 에세이.
065	김영하	나는 나를 파괴할 권리가 있다	1996	M	Canon	신세대 문학의 등장. 죽음의 미학화.
066	최명희	혼불	1996	F	Canon	전통문화와 토속어를 복원한 대하소설.
067	양귀자	모순	1998	F	Popular	인생의 모순을 다룬 대중적 베스트셀러.
068	조창인	가시고기	2000	M	Popular	부성애를 다룬 최후성 베스트셀러.
069	김훈	칼의 노래	2001	M	Canon	이순신을 다룬 역사 소설. 하드보일드 문체.
070	김영하	검은 꽃	2003	M	Canon	멕시코 이민사를 다룬 역사 픽션(Faction).
071	박민규	삼미 슈퍼스타즈의	2003	M	Canon	루저 정서와 인터넷 문화의 결합.

		마지막 팬클럽				
072	한강	채식주의자	2007	F	Canon	맨부커상/노벨상 수상작. 폭력과 식물적 상상력.
073	신경숙	엄마를 부탁해	2008	F	Popular	모성 신드롬을 일으킨 밀리언셀러.
074	유시민	나의 한국현대사	2014	M	Control	대중적 역사 교양서.
075	장강명	한국이 싫어서	2015	M	Popular	'헬조선' 담론과 청년 세대의 이민.
076	정세랑	보건교사 안은영	2015	F	Popular	명량한 히어로물. 20-30대 여성 팬덤.
077	조남주	82년생 김지영	2016	F	Canon	페미니즘 리부트의 상징. 르포르타주 형식.
078	정유정	종의 기원	2016	F	Popular	악의 내면을 다룬 스릴러.
079	손원평	아몬드	2017	F	Popular	감정을 못 느끼는 소년의 성장담 (YA소설).
080	김초엽	우리가 빛의 속도로 갈 수 없다면	2019	F	Popular	한국 SF의 주류 부상. 과학과 서정의 결합.

#### 4. 주석 작업 과정 및 현황

2026년 2월 1일 기준으로 41작품(656 조각)에 대한 샘플링이 완료되었고, 총 1280조각을 기준으로 평균 20.8% 만큼 주석 작업이 완료되었다. 주석 작업은 직접 제작한 전용 웹페이지에 주석자들이 원격으로 접속하여 수행한다. 미리 가공하여 서버에 저장된 텍스트조각 하나씩을 웹페이지에서 제공하고, 주석자는 해당 문장을 읽고 서술 시간을 분 단위<sup>4</sup>로 평가한다.

포스터에 명시한 것과 같이, 세 주석자가 동시에 작업을 완료한 140개 조각에 대하여  $r = .5 \sim .64$ 의 상관관계수 값이 산출되고, 주석 값이 유효함을 알 수 있다.

#### 5. fic-time 라이브러리

현재 영문 텍스트에 대해서 동작하는 fic-time 라이브러리의 초기 버전이 제작되어 있다. 추후 데이터셋이 완성되고 한국어 성능이 검증되면, 한국어 지원을 추가하여 pip 등의 저장소에 배포할 예정이다.

#### 6. 기대효과

본 연구는 'Figure 3'과 같은 절차로 진행되며, 최종적으로 '서술 시간이 기록된 한국어 텍스트 데이터셋'과, '서술 시간을 측정할 수 있는 연구 도구'를 산출한다.

본 연구가 산출하는 1280개 조각의 텍스트는 총 640,000자 분량으로, 한국 문학의 역사적 흐름을 원거리 읽기에 기반하여 조망하기에 충분한 크기를 갖추었다. 이를 통해 다양한 문학사적 연구가 촉발되리라 기대한다.

본 연구는 산출한 데이터셋을 통해, 한국어 소설의 서술 시간을 LLM이 추정할 수 있음을 증명한다. 막연한 직관을 넘어, 실제로 LLM을 통해 서술 시간을 추정할 수 있음이 증명된다면, 문학 작품의 특징과 구조를 분석할 새로운 도구로서 LLM을 사용할 수 있게 된다. 이를 통해 대량의 문학 작품을 서술 시간의 관점에서 연구할 수 있게 될 것이다.

<sup>4</sup> Underwood(2018) 연구는 서술 시간을 분(Minutes) 단위로 측정해야 한다고 밝힌다.