

한국어 과거시제 표지의 병렬 말뭉치 기반 분석*

송상헌

(난양이공대학교)

Song, Sanghoun. 2015. **A Parallel Corpus-based Analysis of the Past Tense Morpheme in Korean**. *Language Information*. Volume 20. 75-104. In human language, a single linguistic form can be used to convey different meanings. One of the most representative forms which involve such a relation between form and meaning in Korean is the verbal inflectional morpheme *(e/a)ss*. This morpheme is responsible for the past tense by default, but in more than a few cases it does not necessarily denote an event that happened in the past. This article probes into the past tense morpheme in Korean *(e/a)ss* in a way of comparative corpus linguistics. In order to create the findings using a data-based method, the present study explores a bilingual parallel corpus in which a sentence in one language is aligned to the corresponding sentence in the other language. The parallel data the current work makes use of is the *Sejong* English-Korean Bilingual Corpus. Exploring the parallel data, this corpus study provides a quantitative analysis of (i) which linguistic form in English the past tense morpheme in Korean corresponds to and (ii) which verbs are more frequently associated with the non-past meanings when they are combined with *(e/a)ss*.

Key words: the past tense morpheme, *(e/a)ss*, tense, aspect, the *Sejong* English-Korean Bilingual Corpus, comparative corpus linguistics

* 세 분의 익명 심사자의 귀중한 지적에 힘입어 본 논문을 보다 정제할 수 있었다. 본고에 남아있는 분석 및 기술상의 오류는 전적으로 필자의 책임임을 밝힌다.

1. 서론

우리는 자연언어에서 어느 하나의 형태가 때로 둘 이상의 의미에 대응하는 경우를 흔히 관찰할 수 있다. 다시 말해, 형태와 의미는 반드시 일대일 관계로 맺어지는 것이 아니다. 한국어에서 이러한 현상을 수반하는 대표적인 사례로 거론되는 것이 이른바 과거시제 표지 ‘있/았’이다.

실제로 시제(tense), 상(aspect), 법(mood) 등을 나타내는 표지의 의미적 기능이 단일하지 않다는 점은 여러 언어에서 관찰되는 바이다. 예컨대, 영어에서 과거시제 표지는 때로 가정법(subjunctive)을 나타내기 위해 사용되며,¹⁾ 이때의 문의 실제 시제는 과거 또는 대과거가 아니다(Iatridou, 2000). 아래 (1a)와 (1c)는 형태적으로는 과거시제 표지가 사용되었을지라도, 그 의미적 시제는 현재이며, 마찬가지로 (1b)와 (1d)는 대과거 표지와는 달리, 단순 과거 사건에 대한 반사실적 명제를 표상한다.

- (1) a. If he were smart, he would be rich.
 b. If he had been smart, he would have been rich.
 c. I wish I had a car.
 d. I wish I had a car when I was a student. (Iatridou, 2000)

이와 같이 과거시제 표지가 때로 다른 시제의 의미를 지니거나 상 또는 법에 해당하는 용법으로 전용되는 경우는 한국어에서도 관찰된다.²⁾ 이러한 반례는 종종 한국어 형태소 ‘있/았’이 과연 시제표지인가에 대한 언어학적 논쟁을 낳게 하는 근거가 되어 왔다. 몇 가지 사례를 보면 아래와 같다.

- (2) a. (전화벨이 울릴 때 상대방에게 전화를 건네주며) 전화 **왔**어요.
 b. (접시를 깨고 나서) 나는 내일 엄마 오면 죽**었**다.
 c. 지금 미아가 부산에 **갔**다. (Lee, 2008)
 d. 네가 지금 여기 **없**으면, 나는 주차도 못 할 거야. (Song, 2014)
 e. 그 놈 참 잘 생**겼**(/*기)다.
 f. 길수는 어머니를 닮**(았)**다. (박동열, 2003)

(2a)에서 전화가 온 상황은 현재 일어나고 있는 일이다. 이 문장을 영어로

1) ‘가정법’이라는 명칭이 때로 혼선을 야기하기 때문에 적절치 않은 문법용어라는 점은 여러 학자들에 의해 지적된 바 있다. 이에 동의하는 바이나, 그 엄밀한 구분이 본고의 주된 관심 사항이 아닌 만큼 여기에서는 일반적인 용어로서 ‘가정법’을 사용하기로 한다.

2) “한국어를 비롯한 많은 언어에서, 하나의 문법요소가 시제, 상, 양태, 증거성, 의외성 중 둘 이상의 범주에 걸쳐서 복수의 의미 성분을 갖는 일이 흔히 있다.” (박진호, 2011:289)

번역하였을 경우 과거시제 표지 ‘-ed’가 사용되면 문맥상 어색한 표현이 될 것이다. (2b)는 보다 직접적인 반례로서 ‘있’이 사용되었으나 해당 문장은 미래에 일어날 상황에 대한 추측에 해당한다. (2c)의 예문에서는 과거시제 표지가 현재 사태를 나타내는 부사 ‘지금’과 같은 절에서 공기하고 있다. 마찬가지로 과거시제 표지가 ‘지금’과 같은 절 안에서 공기하고 있는 (2d)의 예문은 한국어에서도 때에 따라 과거시제 표지가 반사실적 명제를 표시하는 기능을 수행하고 이때 실제 의미상의 시제는 현재일 수 있다는 점을 보인다. (2e-f)의 예문은 보다 특징적이다. ‘생기다’, ‘답다’와 같은 특정 동사의 경우 과거시제 표지를 수반할 경우에만 정문으로 판별될 수 있으며, 이때의 문장은 과거의 사건이 아닌 현재의 상태를 기술한다.

이러한 일련의 반례에도 불구하고 한국어학 일반에서는 ‘있/았’을 시제표지로 파악하는 시각이 더 우세한 듯 보인다.³⁾ 대부분의 한국어 기본 문법서는 한국어의 시제체계 안에 과거시제를 포함시키고 있으며, 해당 예문은 ‘있/았’을 수반하여 기술된다(장석진 1995; Lee and Ramsey 2000; 김종복 2004; Sohn 2013). 아마도 가장 큰 이유는 ‘있/았’이 과거시제 표지로 기능하는 것이 보다 빈번한 확률이라고 가정되기 때문일 것이다. 문제는 이러한 예상이 실제 자료를 통해 신뢰성 있는 수준으로 검증된 바가 없다는 것이다.

이러한 맥락에서 본고는 다음의 질문에 대한 답을 얻고자 한다. 첫째, 한국어의 ‘있/았’은 어느 정도의 비율로 과거시제를 표현하는가? 둘째, 한국어의 ‘있/았’이 과거시제를 표현하는 기능을 수행하지 않을 때 ‘있/았’과 결합하는 서술어에는 어떠한 속성이 발견되는가? 본고는 말뭉치 언어학의 방법을 통해 이를 밝히고자 한다. 실제 언중이 사용하는 자료를 통해 한국어의 ‘있/았’이 수행하는 기능에 어떠한 의미적 범주들이 포함되어 있는가를 밝히고, ‘있/았’이 각 의미적 속성에 대응하는 비율이 실제로 어느 정도로 분포하는가를 파악할 것이다. 여기에서 제기되는 문제점은 각 의미적 범주에 대한 주석을 어떠한 일관되고 객관적인 준거에 따라 기술하는가이다. 실제 위 (2)에서 제시된 예문에서 본 바와 같이 ‘있/았’의 의미적 속성이 무엇인가를 정확하게 설명하기가 모호한 실례가 더러 있다. 이에 대한 판별 근거를 미리 확립해 놓지 않으면, 그 분석 결과 역시 연구자의 주관 또는 이론적 배경에 따라 편향성을 지니게 마련이다. 자료중심의 객관적 판단을 기하기 위해 본 연구에서는 한영 병렬 말뭉치를 사용하고자 한다. 병렬 말뭉치는 복수의 언어의 문장을 문장 대

3) ‘시제’라는 용어가 관점에 따라 혼선을 야기할 수 있음은 주지의 사실이다. 한 심사자는 본고에서 ‘과거시제’라고 칭한 상당수가 실제로 ‘과거사태’를 가리키고 있음을 지적하였다. 해석론적 관점에서 보면 이는 타당한 지적인데, 다만 본고에서 사용되는 ‘과거시제’라는 명칭은 표현론의 관점을 따른 것임을 밝힌다.

문장 또는 단락 대 단락으로 나란히 주석 처리해 놓은 텍스트를 말하는데, 최근의 대조언어학적 관점의 말뭉치 연구에서 자주 사용되는 언어 자원이다.

본고의 구성은 다음과 같다. 2절은 한국어 표지 ‘있/았’에 대한 선행연구를 간략히 개괄하고, 본 말뭉치 연구의 기본 방향을 정한다. 3절은 해당 연구에 말뭉치적 접근법이 필요한 이유를 구체적으로 기술하고, 본 연구의 수행을 위한 연구 방법론을 제시한다. 4절은 그 방법론에 입각하여 세종 한영 병렬 말뭉치(서상규 외, 2007)에서 습득된 결과를 제시한다. 5절은 그러한 분포적 결과를 낳는 언어적 요인이 무엇인가를 추론하기 위한 계량 분석을 실시한다. 6절은 본고의 결론이다.

2. 논의의 배경

앞 절에서 언급한 한국어 문법서 이외에도, 한국어 ‘있/았’에 관련된 개별 선행 연구는 일일이 나열하기가 불가능할 정도로 많다. 연구의 관점에 따라 대표적인 것만을 추려보면, 이론적 차원의 연구(아소요코 2004, 문숙영 2005;2008, 박진호 2011, 등), 대조언어학적 연구(황원미 2003, 박동열 2003, 김현식 2004, 등), 시간 정보 태깅에 관한 연구(Lee 2008, 유현조 외 2011, 김문형 외 2012), 실험 및 심리언어학 연구(박현아·오은정 2014), 언어습득 연구(강남욱 2011, 김창구 2012, 장기남 2013, 등) 등이 망라된다. 본고의 주된 목적은 ‘있/았’과 관련된 말뭉치 주석처리와 이에 근간한 계량적 분석인 바, 이어지는 각 소절에서는 말뭉치의 분석과 직접적으로 관련된 사항을 선별적으로 짚어보고자 한다.

2.1. ‘있/았’의 문법적 지위

앞 절에서 언급한 바와 같이, 한국어의 ‘있/았’은 대체로 과거시제를 나타내는 표지로 이해되어 왔다. 그러나 이에 대한 이론적 반론 역시 적지 않다. 예컨대, ‘있/았’은 시제 표지라기보다는 상 표지라고 보는 입장에서부터 나아가 과거시제가 한국어에서 실제하는가에 대한 근본적 문제제기도 존재한다. 아소요코(2004)는 한국어 형태소의 ‘있/았’의 문법적 지위와 관련한 선행 연구를 크게 네 가지 입장으로 정리하였다.

- (3) a. 시제 범주로 다루면서 ‘-었-’이 ‘과거’의 의미를 가진다고 보는 연구
- b. 상 범주로 다루면서 ‘-었-’이 ‘완료’의 의미를 가진다고 보는 연구
- c. 시제 범주와 상 범주를 나누지 않고 ‘-었-’을 ‘과거’와 ‘완료’의 의미로 보는 연구
- d. 서법 범주로 다루면서 ‘-었-’이 ‘완결’, ‘사건의 전체성에 대한 인식의 종결’ 등 사건에 대한 화자의 태도를 나타낸다고 보는 연구

이 가운데서 본고가 ‘있었’의 문법적 지위에 대하여 취하는 입장은 ‘있었’의 가장 일차적인 기능을 어떠한 사건 혹은 어떠한 상태가 과거의 일이었음을 나타내는 것으로 파악하는 박진호(2011) 및 박현아·오은정(2014)와 동일하다. 박진호(2011)은 “과거 사태를 표현할 때 과거 표지가 필수적으로 사용되는가를 표현론적으로 살펴보아야 한다”고 주장하며, 이러한 관점에서 한국어에서 ‘있었’은 분명한 과거시제 표지라고 밝히고 있다. 이는 ‘있었’을 기본적으로 과거시제 표지로 인정하면서, 이에 더하여 다른 용법이 존재하는 것으로 그 문법적 지위를 판단하는 것이다.

이에 대한 반대 의견 역시 존재한다. (3b)의 관점에서 양용준(2002)은 ‘있었’의 속성과 관련하여 상 범주만을 논의의 대상으로 한다. 예컨대, 아래 인용된 예문에서 ‘있었’을 상을 표현하는 범주에 포함시켰다.⁴⁾

- (4) a. 폭풍우가 몰아치면서 번개가 번쩍하였다. (순간상)
- b. 나의 누나가 인형을 다 만들었다. (종결상)

그러나 필자의 직관으로는 위 두 예문에서 ‘순간상’ 또는 ‘종결상’을 낳는 기제는 ‘있었’이라기 보다는 밑줄 친 어휘들(‘번쩍하-’, ‘다’)이다. 다른 한편으로 (3c)의 입장을 견지하는 박영준(1998)과 박동열(2003)이 있다. 이는 ‘있었’을 과거의 ‘있었’과 완료의 ‘있었’으로 나누어 보는 것인데, 특히 박영준(1998)은 통시적 관점에서 한국어의 ‘있었’을 세 가지 다른 형태로 나누어 기술한다. 다시 말해, 한국어에는 ‘있었1’, ‘있었2’, ‘있었3’가 별도의 어휘항으로 존재한다는 관점이다. 물론, 역사적인 어휘 변천의 차원에서 과거 의미가 강한 ‘있었’과 완료 의미가 강한 ‘있었’의 유래가 다를 수는 있겠으나, 공시적 차원에서 보면 이들의 분포적 양상에는 큰 차이가 없다.⁵⁾ 통시적 변화에 대한 관찰이

4) 물론 양용준(2002)는 한국어의 경우 영어보다 어휘상을 정확히 판별하기 어렵고, 한국어에서 상 표지는 영어보다 훨씬 다양하다는 점을 지적하고 있다.

5) 이에 대한 계량적 증거는 4절에서 제시된다.

현대 한국어의 쓰임을 분석하는 데 있어 요긴한 도구인 것은 분명하나, 현대 한국어의 체계까지 결정짓는 절대적 기준은 아니라고 여긴다.⁶⁾

정리하자면 한국어에서 ‘있/았’의 가장 기본적이고 대표적인 쓰임은 과거시제 표지이기 때문에 이 쓰임을 중심으로 ‘있/았’을 이해하는 것이 자연스러운 접근법이다. 본고는 더 나아가 ‘있/았’의 쓰임을 논리식으로 전환하여, ‘있/았’은 과거 시제를 표현하는 ‘필수조건’이라고 가정한다. 즉, ‘있/았’이 사용되었다고 해서 모두 과거일 수는 없으나, 과거 사태를 표현하고자 한다면 필수적으로 ‘있/았’이 출현해야 한다는 것이다.⁷⁾ 바꾸어 말하면, ‘있/았’이 담당하는 기능 가운데 ‘과거’는 부분집합을 이루며, 동시에 가장 큰 부분집합을 이룬다.

‘있/았’이 과거 사태 표현의 필수조건이라는 주장에 대한 반론도 있을 수 있다. 예컨대, 문숙영(2008)은 어떠한 명제가 발화시 이전에 발생한 과거 상황이라 하더라도 이 명제에 반드시 ‘있/았’이 사용되어야할 강제성은 없다고 주장한다. 문숙영(2008)은 그 대표적 근거로 보도문에서 추출한 아래와 같은 예문을 들었다.

- (5) 1868년 일본 메이지정부는 울릉도와 독도가 조선의 부속 영토로 된 경위를 조사하라고 지시합니다.

문숙영(2008)은 위 (5)와 같은 예문을 “역사적 현재의 예”라고 칭하였다. 그러나 필자는 위 (5)에서 과거시제 표지가 사용되지 않은 이유를 보도문의 장르적 특성 때문이라고 본다. 즉, 보다 생생한 사건 서술을 위하여, 화자가 그 장면 속으로 들어가서 사건을 현재형으로 기술하는 문체를 사용하는 것이다. 이와 유사한 예는 시제의 표현이 한국어보다 엄격한 영어에서도 종종 관찰되는 현상이다. 아래 (6)은 영어에서 시제 표현이 그 사태의 시간성과 동일하지 않은 예이다.

- (6) a. She comes here the day after tomorrow.
 b. If it rains tomorrow, the picnic will be called off.
 c. He argued that the honesty is the best policy.
 d. We learned that the light travels faster than the sound.

6) 보다 본질적인 차원에서의 문제제기도 있을 수 있을 것이나 이에 대한 반론은 박진호(2011)의 주장으로 같음하고자 한다. ‘있/았’이 항상 과거 사태에만 사용되는 것은 아니기 때문에, ‘있/았’을 과거시제 표지로 상정할 수 없다는 주장(이효상 1995)은 주변부의 용법을 근거로 하여 기본적인 용법을 무시하는 관점이며, “한국어는 무시제 언어”라는 견해(남기심 1978)는 언어 분석에 있어서 표현론적 접근을 도외시하였다는 측면에서 타당성있는 주장이 아니라는 것이다.

7) 그러나 이와 별도로 실제 자료를 관찰하였을 때, 앞서 제시한 논리관계에 대한 반례가 발견되었는데, 이는 3.4절에서 구체적인 예시와 함께 언급될 것이다.

‘있/았’이 과거 시제를 위한 필수조건이라는 주장은 다른 한편으로 ‘있/았’이 출현하였다 하더라도 해당 명제가 반드시 과거가 아닐 수 있다는 점을 함의한다. 이에 대한 예시는 앞 (2)에서 이미 제시되었다. 이러한 차이를 낳는 원인은 무엇인가? 이에 대한 이론적 설명은 아소요꼬(2004)와 Lee(2008)에서 일부 제시되고 있다. 편의상 몇몇 예문을 아래에 다시 기술하기로 한다.

- (7) a. (접시를 깨고 나서) 나는 내일 엄마 오면 죽었다.
 b. 지금 미아가 부산에 갔다.
 c. 길수는 어머니를 닮았다.

먼저 (7a)와 관련하여 아소요꼬(2004)의 주장을 따르면, 해당 문장에서 과거 시제 표지가 사용된 이유는 해당 명제가 화자의 강한 믿음을 반영하기 때문이라는 것이다. 즉, 마치 과거에 있었던 일인 것처럼 해당 사건이 반드시 발생할 것이라는 예측을 나타내는 용법이다. 다음으로 (7b)의 문장에서 ‘있/았’이 사용된 이유에 관련하여 Lee(2008)의 주장은 해당 사건이 과거의 특정 시점에 발생하였으며 그 사건이 현재까지도 지속되기 때문이라는 것이다. 이는 영어의 표현론으로는 현재완료에 해당하는 용법이다. 이러한 관점에서 보면 (7c)에서 과거시제가 필수적으로 사용된 이유도 일부 설명할 수 있는데 과거부터 현재까지의 지속적인 상태를 나타내기 때문이다. 위 세 문장 모두 엄밀한 의미의 과거는 아니지만, 그 의미해석에 있어서 과거와 관련된 해석이 전혀 없다고는 볼 수 없다. 다시 말해, 그 의미가 이루어지는 과정에 있어서, 과거의 사태를 참조해야할 필요성이 내재된 논리 구조이다.

2.2. ‘있/았’의 중복 사용

한국어에서 ‘있/았’과 함께 과거시제를 표시하는 표지로 거론되는 형태로 ‘있/았’의 중복사용이 있다. 문숙영(2008)에서 제시된 예문은 아래와 같다. 문숙영(2008)의 주장에 따르면 이는 사건이 역순으로 배열됨으로써 ‘있/았’의 대과거성이 드러나는 예이다.

- (8) a. 당시 내무부는 감사원의 압행감찰에서 적발된 당시 의정부 시장을 직위 해제하고 해임을 요구했었습니다.
 b. 경총은 올해 임금인상 가이드라인으로 3.9%를 제시했었습니다.

본고에서는 크게 세 가지 이유에서 중복 사용된 ‘있/았’의 형태를 분석의

대상에 포함시키지 않는다. 첫째, 이때 ‘있/았’이 중복해서 사용되었다 할지라도 이 표지가 영어의 대과거와 의미·화용적으로 동일한 것은 아니다 (장석진 1995, Lee 2008). 둘째, 실제 자료에서 ‘-있었.’이 기대될 만한 위치에 단일 ‘있/았’이 사용되는 예가 다수 발견된다는 문숙영(2008)의 관찰이 존재한다. 즉, 한국어의 ‘있/았’의 중복 사용은 엄격한 규칙에 의해 생산되는 형태가 아니다. 끝으로 ‘-있었.’의 상정이 완료의 의미를 파악하는 언어실험에 있어서 그다지 유용하지 않았다는 박현아·오은정(2014)의 경험적 관찰도 ‘있/았’의 중복 사용을 자료 분석에서 배제하는 근거가 된다.

2.3. 문법상과 어휘상

이른바 상이라는 범주는 크게 문법적인 상과 어휘적인 상으로 구분된다. 이는 비단 상적 범주뿐만 아니라 시제 범주 등에 있어서도 마찬가지로 적용된다. 한 언어에서 문법적 범주는 그 수가 다소 제한적인 반면 어휘적 범주는 언제나 생산적인 특성을 지닌다. 예컨대 영어에서 문법적 상은 크게 ‘have+pp’와 ‘be+V-ing’의 두 가지 형태로 실현되지만, 어휘적 상은 동사의 속성에 따라 다양하게 나타날 수 있다. 한국어의 ‘있/았’을 분석하는 과정에 있어서도 이러한 구분은 고려되어야 할 사항이다. 본고에서는 앞 2.1절에서 언급한 바와 같이 ‘있/았’을 기본적으로 과거시제 표지로 파악하며, ‘있/았’이 포함된 문장이 상적 의미를 가지는 경우 이는 동사(또는 다른 단어)에서 파생된 어휘상으로 파악하고자 한다. 따라서 ‘있/았’의 대응관계를 조사함과 동시에 ‘있/았’이 어떠한 동사와 결합하는가를 살피고자 한다.

2.4. 요약

이상의 내용을 요약하면 다음과 같다. 우선 본고는 ‘있/았’의 기본 기능을 과거시제 표지로 파악하고 이후 일관되게 ‘있/았’을 과거시제 표지로 칭할 것이다. 아울러 ‘있/았’은 과거 사태를 표현하는 필수조건으로 가정한다. ‘있/았’의 중복 사용은 실제 자료 검토에서 유용한 정보가 아닐 가능성이 크기 때문에 분석의 대상에서 제외한다. 끝으로 ‘있/았’과 결합하는 동사의 상적 속성이 중요하다라는 점에서 어떠한 동사가 ‘있/았’과 결합하였을 때 과거 의미를 더 유의미하게 취하는지를 관찰할 것이다.

3. 방법론 설정

본고는 한국어의 ‘있/았’의 쓰임을 밝히기 위해 말뭉치 기반 접근을 취한다. 한국어의 표지 ‘있/았’의 의미적 기능에 대한 기존의 연구는 대체로 연구자가 직접 구성한 예문을 분석한 이론적 연구에 속한다. 물론 이러한 방식의 연구 자체가 잘못된 것은 아니며 오히려 ‘있/았’의 문법적 지위에 관련한 다양한 설명을 시도하였다는 점에서 많은 시사점을 지닌다. 그러나 기존의 연구 방식에서 한계로 지적될 수 있는 부분 역시 존재한다. 첫째, ‘있/았’의 문법적 쓰임을 연구자마다 상이한 직관과 이론적 배경에 따라 분석한 것이기 때문에 때로 특정 이론의 틀에 치우친 논의가 발생할 수 있다. 둘째, 다양한 가능성을 포함한 실제 자료를 분석한 것이 아니기 때문에 그 연구 결과가 한국어에서 ‘있/았’이 차지하는 문법적 속성을 어느 정도 포괄하고 있는지에 대한 전체 그림을 한눈에 파악하기가 쉽지 않다.

기존의 이론적 연구와 병행하여 말뭉치 분석을 도모하는 것은 한국어의 ‘있/았’에 대한 보다 종합적이고 균형잡힌 연구 성과를 도모하기 위한 중요한 기제이다. 실제 언증에 의해서 생산된 자료의 모음인 말뭉치에 근거하는 연구 방식은 연구 결과의 객관성을 향상시켜주는 주효한 방법임에 대해서는 언어학 연구의 최근 흐름에 비추어 이견이 없을 것이다. 아울러 말뭉치 분석은 상당한 크기의 자료를 통해 계량적으로 입증된 결과를 제시하는 바, 기존 연구의 주장이 어느 정도의 타당성이 있는가에 대한 수치적 검토를 가능케 한다.

3.1. 선행 연구

실제 언어 자료를 통해 ‘있/았’의 쓰임을 분석한 선행 연구가 전무한 것은 물론 아니다. 대표적으로 앞서 언급한 문숙영(2008)은 보도문을 대상으로 하여 ‘있/았’을 비롯한 시제 표지의 용례를 분석하였다. 이 연구는 이론적 기반에서 다양한 시제 표지를 면밀히 관찰하였다는 점에서 가치를 가지고 있다. 그러나 그 분석의 대상이 보도문이라는 특정 장르로 국한되어 한국어의 실제 쓰임에 관한 전체적인 양상을 다 보여주지는 못한다는 점과 자료를 분석하는 과정에서 계량적 분석이 다소 제한적이라는 점에서 일정 정도의 한계를 지닌다.

본고와 마찬가지로 번역문 다시 말해 병렬 텍스트를 기반으로 하여 한국어 ‘있/았’의 쓰임을 비교분석한 연구로는 황원미(2003)을 들 수 있다. 이는 프랑스어가 원전인 소설의 한국어 번역본을 대상으로 하여, 프랑스어에서 각 시제가

한국어 ‘있/았’과 어떠한 대응관계를 보이는지 분석한 연구이다. 전체 ‘있/았’의 출현빈도는 총 3,323회로 집계되었으며, 한국어의 ‘있/았’이 프랑스어에 대응하는 양상은 반과거 20.13%, 복합과거 35.6%, 대과거 3.58%, 단순과거 0.18%로 집계되었다. 이는 이미 출판되어 있는 공개 자료를 통해 입증을 하였다는 점과 시제의 체계가 세분화된 프랑스어를 비교의 대상으로 삼았다는 점에서 참고의 가치가 높은 연구이다. 그러나 ‘있/았’과 결합된 동사들을 주요하게 관찰하면서도 그 동사들의 특성에 대한 계량적 분석에까지는 이르지 않았다.

다른 언어의 예로서는 趙世開·沈家煊(1984)를 참조할 수 있다. 이는 중국어의 了를 대상으로 하여 위 황원미(2003)과 유사한 방식의 분석을 진행한 결과이다. 다만, 참조의 대상이 되는 언어가 영어라는 점에서 차이가 있다. 중국어에서 일반적으로 了는 보통 완료상의 표지로 여겨지는데, 了는 동사에 바로 결합하는 동사 불변화사 了1과 문장의 말미에 출현하는 了2로 구분된다. 이러한 구분에 따른 비교 결과 趙世開·沈家煊(1984)는 了1과 了2 모두 영어의 과거 시제에 대응하는 경우가 가장 많으며(전체의 52.42%), 완료상 표지로 간주되던 了는 실제 영어의 완료상과 상당히 다른 분포를 보인다는 점을 발견하였다. 了가 사용된 경우 대응하는 영어 표현이 완료상으로 실현된 경우는 전체의 22.95%에 불과하였다.

3.2. 말뭉치 분석의 준거

언어자료를 계량적으로 분석하기 위해서는 다시 몇 가지 측면이 세부적으로 고려되어야 하는데, 말뭉치 분석과 관련된 기본 원칙은 다음과 같다.

무엇보다 먼저, 자료를 분석하는 과정에 있어서 객관적인 판별기준이 마련되어야 한다. 말뭉치 언어학의 가장 큰 장점은 객관화된 연구 결과를 도모할 발판을 마련하는 데 있을 것이다. 그러나 객관적으로 수집된 자료를 주관적인 기준에 의해 처리한다면 이는 말뭉치 언어학의 장점을 제대로 살리지 못하는 일이다. 따라서 ‘있/았’이 가지는 쓰임이 무엇인가를 미리 설정된 판별 기준에 따라 기계적으로 구분하여 연구자의 주관에 개입할 여지를 최소한도로 줄이는 장치가 필요하다.

둘째, 충분한 양의 자료를 분석의 대상으로 삼는다. 어느 정도의 규모를 살펴보아야 충분한 관찰인가에 대해서는 절대적 기준이 존재하는 것이 아니다. 자료의 충분성은 대체로 분석의 대상이 되는 언어 현상이 무엇인가에 따라 상대적으로 결정된다. 한국어의 ‘있/았’은 그 사용빈도가 높다는 점에서 천만 어절 이상의 자료를 살펴볼 필요까지는 없을 테지만, 합리적 일반화를 위해

최소 몇 십만 어절 이상의 자료가 관찰되어야 할 것이다.

셋째, 다양한 장르의 텍스트가 분석의 대상에 포함되어야 한다. 한국어는 텍스트의 장르에 따른 언어적 속성의 차이가 비교적 크고, 그러한 장르적 속성의 차이가 ‘있/았’의 출현에 있어서 영향을 주지 않는다는 보장이 없다. 따라서 한국어에서 ‘있/았’의 기능이 무엇인가를 종합적으로 살피기 위해서는 균형 말뭉치를 살펴야 한다.

넷째, 연구 결과의 재현성이 보장되어야 한다. 말뭉치 언어학은 계량 학문으로서의 성격을 지니기 때문에, 연구 결과를 재구할 수 있으나 문제가 언제나 함께 고려되어야 한다. 연구 결과를 다시 구성할 수 있도록 하는 가장 **일차적인** 선택은 공개된 언어 자료를 활용하는 것이다. 즉, 누구나 적절한 절차를 거친 이후라면 열람할 수 있는 자료를 분석의 대상으로 삼아야 한다. 더 나아가 본 연구의 결과 일체는 추후 연구자들이 분석하고 사용할 수 있도록 공개될 것이므로, 그 원천 자원 역시 공개 자원이어야 한다.

3.3. 세종 한영 병렬 말뭉치

앞 소절에서 논의된 준거를 모두 충족하기 위해 본 연구에서 분석의 대상으로 삼는 언어 자원은 ‘세종 한영 병렬 말뭉치’이다(서상규 외, 2007). 병렬 말뭉치란 복수의 언어를 대상으로 하여 한 언어의 문장 또는 단락이 다른 언어의 문장 또는 단락과 가지는 대응쌍을 주석처리한 언어 자원을 말한다. 시제표지의 쓰임을 밝히기 위해 병렬 말뭉치를 참조하는 기본 구상은 3.1에서 언급된 趙世開·沈家煊(1984) 및 황원미(2003)과 동일하다.

선행연구와 한 가지 관점상의 차이가 있다면, 趙世開·沈家煊(1984), 황원미(2003) 등은 주로 대조언어학의 견지에서 시제 혹은 상 표지로 알려진 중국어의 了와 한국어의 ‘있/았’이 다른 언어에 어떻게 대응되는지를 검토하였다면, 본 연구에서는 객관적 분석을 위한 장치로서 병렬 말뭉치를 사용한다. 한국어의 ‘있/았’의 경우 이론마다 설명의 방식이 다르고 연구자마다 직관이 상이할 수 있기 때문에, 일정한 판별기준에 근거하지 않고서는 일관된 주석처리 작업이 사실상 불가능하다. 본고에서 한영 병렬 말뭉치를 사용하는 것은 연구자의 주관적 판별을 배제하기 위한 일종의 기계적 장치이다. 즉, 어떻게 표현되었는가를 주석처리의 유일한 기준으로 삼는다.

물론 영어의 시제 체계가 언어 현상 전반을 아우르는 절대적인 기준도 아니며, 영어의 시제 및 상이 한국어와 구조적인 대응관계를 지니는 것도 결코 아니다. 그러나 영어의 시제와 상은 표지가 비교적 뚜렷하기 때문에

이러한 형태적 기준에 따라 주석처리를 한다면 최소한 그 범위 안에 있어서는 보다 객관적인 분석을 도모할 수 있을 것이다. 다시 말해, 한국어의 ‘있/있’이 영어의 대응 문장에서 어떻게 표현되었는가를 참조하여서 그 의미를 규정짓는 절차적 알고리즘을 활용하면, 자료의 주석처리를 하는데 있어 연구자가 개입할 여지를 상당 부분 차단할 수 있다.⁸⁾

관련하여, 영어에서 시제와 상의 표지가 지니는 기능을 검토해 보자. 앞서 한국어에서 ‘있/있’을 과거시제를 표현하기 위한 ‘필수조건’으로 상정하였다. 마찬가지로 표현론적 견지에서 영어의 시제를 살펴보면 영어 말뭉치에서 통상 VBD로 주석처리되는 ‘-ed’ 형태는 과거시제를 표현하는 ‘필요충분조건’에 해당할 것이다. 물론 이 정의에 대한 반례도 존재한다. 예컨대 (1)에서 살펴본 바와 같이 가정법 구문에서는 ‘-ed’의 의미적 시제는 ‘현재’이다. 따라서 보다 정확히 정의하자면, 영어에서 문장이 직설법을 취하고 있을 때 ‘-ed’는 과거시제와 ‘필요충분조건’ 관계를 가진다. 분석의 과정에서 가정법으로 실현된 문장을 대상에서 제외한다면, ‘-ed’와 과거시제는 사실상 일대일 대응관계를 가진다.

병렬 말뭉치를 사용하여 주석처리하는 방법을 사용한다고 하여도 굳이 영어를 그 비교의 대상으로 가정할 필요는 없으며, 또한 영어를 대상으로 한다고 하더라도 세종 한영 병렬 말뭉치를 대상으로 할 필요는 없을 것이다. 그러나 3.2에서 논의된 분석 준거를 모두 만족하는 자원은 현재로서 세종 한영 병렬 말뭉치일 것이다.⁹⁾ 즉, 상당한 규모로 주석처리된 균형 말뭉치라는 점, 그리고 공개된 자원으로서 누구나 접근이 가능하다는 점에서 세종 한영 병렬 말뭉치는 현재로서 가장 유용한 분석 대상이다.¹⁰⁾

8) 한 심사자는 “기계 번역 연구의 관점에서 한국어와 영어의 시제 대응 양상을 밝히는 목적이려면 모르겠으나, 한국어 자체의 특성을 밝히기 위해 병렬 말뭉치를 사용하는 것이 어느 정도의 타당성을 지니는가”에 대한 다소 회의적인 입장을 표명하였다. 물론, 본 연구의 결과의 일차적인 활용대상은 한국어-영어의 기계번역을 위한 전환규칙 생성의 참조자료일 것이다. 그러나 본고에서 병렬 말뭉치를 사용한 또 다른 그리고 더 본질적인 목적은 일관된 구축지침을 사용하기 위해서이다. 필자는 한국어의 시제 체계를 파악하기 위해 병렬 말뭉치를 사용하는 것이 최고의 해법이라고 주장하는 것이 아니다. 본고에서 제시된 방법 이외에도 여타의 주석처리 방법론이 존재할 수 있다. 다만 특정 이론 또는 특정 직관에 입각하여 주석처리가 이루어지지 않도록, ‘표현되는바’에만 집중하기 위한 선택으로 병렬 말뭉치를 사용한 것이다.

9) 공개된 자원으로서 세종 한일 병렬 말뭉치(서상규 외, 2007)를 사용할 수도 있을 것이나, 일본어의 시제 표지 ㅆ는 한국어의 ‘있/있’과 마찬가지로 중의적이다.

10) 세종 한영 병렬 말뭉치의 구성 및 특성에 관련하여서는 서상규 외 (2007) 또는 Song (2014)를 참조 바란다.

3.4. 주석처리의 방법

한국어의 ‘있/았’의 형태 탐색을 위하여 본 연구에서는 세종 한영 병렬 말뭉치 가운데 형태소 분석이 이루어진 버전을 사용한다. 병렬 말뭉치는 문장 또는 단락 단위로 대응이 이루어지는데, 세종 한영 병렬 말뭉치는 엄밀한 의미에서 문장 단위의 대응이 아니기 때문에 자료 분석의 첫 번째 과정은 문장 단위에 가까운 형태로 원자료를 재구성하는 일이다. 이 과정은 Song(2014)에서 이미 구축한 중간 결과를 재활용하였다. 예시 문장대응셋은 아래와 같다. 먼저 형태소 분석이 이루어진 한국어 문장이 제시되고 다음 라인에 주석처리를 위한 중간행(‘-’로 시작됨)이 위치한다. 그에 이어 해당 형태소 단위로 분석된 영어 문장이 뒤따르는 형식이다.

(9) 연락/NNG+을/JKO 기다리/VV+고/EC 있/VX+겠/EP+습니다/EF+./SF

I/PRP 'll/MD look/VB forward/RB to/TO hearing/VBG from/IN you/PRP ./.

본 연구에서 활용된 자료는 총 38,804개의 문장대응셋으로 구성되어 있으며, 그중 한국어 부분은 총 어절 279,677개, 총 형태소 622,881개로 이루어져 있다.

세종 형태분석 말뭉치에서 과거시제 표지 ‘있/았’은 크게 몇 가지 이형태로 태깅되어 있는데, 여기에는 ‘쓰/EP’ (6,156회 출현), ‘있/EP’ (3,128회 출현), ‘았/EP’ (1,096회 출현), ‘였/EP’ (458회 출현)가 포함된다. 총 10,838회 출현하는 것으로 이는 전체 출현 형태소의 약 1.74%에 해당한다. 아울러 ‘있/았’의 이형태가 하나 이상 출현하는 문장의 개수는 9,357로 집계되었는데, 이는 전체 문장대응셋의 24.11%에 해당한다.¹¹⁾

다음의 과정은 영어 문장이 가정법으로 실현된 문장대응셋을 제외하는 일이다. 이전 소절에서 언급한 바와 같이 가정법 문장에서는 영어의 ‘-ed’가 과거시제와 일대일 대응관계를 이루지 않기 때문에 오분석이 없도록 가정법이 나타난 문장셋을 분석 대상에서 배제시켜야한다. 이 과정은 Song(2014)의 추출결과를 활용하였다. Song(2014)에서는 같은 자료를 대상으로 영어에서 가정법이 나타난 문장셋 139개를 정리하였는데, 이들은 본 연구의 분석 대상에서 일괄 삭제되었다.

이상의 예비 분석 단계가 정리되고 난 이후, 본격적인 주석처리 작업이 이루어졌다. 앞서 언급한 ‘있/았’의 복수 이형태를 검색하기 위해, 아래 (10)의

11) 이 수치는 이후 더 큰 규모의 자료인 세종 형태분석 말뭉치 등을 통해 재검토되어야 한다.

정규표현식을 활용하였으며, 정규표현식 검색은 grep명령어를 활용하였다.

(10) [*(았었였)/^+

위 정규표현식을 통해 ‘았/었’의 형태가 검색되면, 다음으로 해당 영어 문장에서 대응되는 표현의 시제와 상을 참조하였다. 대응되는 영어 표현의 시제와 상이 무엇으로 실현되었는지를 확인하여 그 정보를 ‘_’로 시작하는 중간행에 기입하는 방식을 택하였다. 예컨대, 해당 영어 표현이 VBD로 태깅되었고, ‘have+p.p’의 구성이 아니라면, ‘과거’가 중간행에 삽입되고, 해당 영어 표현이 VBD로 태깅된 단어 없이 ‘have+p.p’로만 구성되었다면 ‘현재완료’로 주석처리되었다. 몇 가지 예시는 아래와 같다.

- (11) a. 예/IC+./SP 어제/NNG 오후/NNG 얘기/NNG+를/JKO
하/VV+ ㄷ *(EP+습니다/EF+./SF
--- 과거
Yes/UH ./, I/PRP talked/VBD him/PRP yesterday/NN afternoon/NN ./.
- b. 기다리/VV+고/EC 있/VX+였/EP+는데요/EF+./SF
--- 과거진행
We/PRP were/VBD waiting/VBG for/IN you/PRP ./.
- c. 어떻게/MAG 지내/VV+ ㄷ *(EP+습니까/EF+?/SF
--- 현재완료
How/WRB have/VB you/PRP been/VBN ?/.
- d. 오/VV+시/EP+ ㄷ *(EP+다고/EC 말씀/NNG+드리/VV+죠/EC+./SP
브라운/NNP+씨/XSN+./SF
--- 현재
I/PRP 'll/MD tell/VB him/PRP you/PRP are/VBP here/RB ./,
Mr./NNP Brown/NNP ./.

하나의 문장에서 ‘았/었’은 경우에 따라 여럿 출현할 수 있는데, 여기에는 두 가지 가능성이 있다. 하나는 ‘았/었’이 하나의 어절 안에서 중복 출현하는 경우이다. 이 경우, ‘았/었’의 중복출현을 분석의 과정에서 포함시키지 않는다는 원칙에 따라 추가적 주석처리를 하지 않았다. 다른 경우로 ‘았/었’이 각기 다른 어절에 출현하는 경우가 있을 수 있는데, 이 경우 중간행에 각각의 정보를 순차적으로 기입하고 ‘/’를 구분자로 사용하였다. 예시는 아래와 같다.

- (12) 들어오/VV+ㅏ ㅁ/EP+다가/EC 몇/MM+분/NNB 후/NNG+에/JKB
 다시/MAG 나가/VV+ㅏ ㅁ/EP+어요/EF+./SF
 --- 과거 / 과거
 She/PRP came/VBD and/CC went/VBD in/IN a/DT few/JJ minutes/NNS ./.

경우에 따라 영어에서 대응하는 영어 표현이 VBD태그를 달고 있지 않더라도 과거 시제가 표현된 것으로 판단해야 할 경우가 있는데, 대표적으로 MD로 태깅된 서법조동사가 사용된 경우이다. 예시는 아래와 같다.

- (13) 273/SN 도저히/MAG 자리/NNG+를/JKO 뜨/VV+ㄹ/ETM
 수/NNB+가 /JKS 없/VA+였/EP+어요/EF+./SF
 --- 과거
 273/CD I/PRP just/RB couldn't/MD get/VB away/RB ./.

영어에서 would, could 등의 단어는 will 또는 can의 과거시제 표현으로 기능을 할 수도 있으나, 반드시 그러한 것은 아니다. 세종 한영 병렬 말뭉치는 문맥을 살필 수가 있도록 구성되어 있기 때문에, 이와 같은 경우에는 앞뒤 문맥을 참조하였다.

정규적 대응관계 이외에도 몇 가지 주변적인 대응관계가 또한 발견되었다. 첫째, 세종 한영 병렬 말뭉치는 실제 텍스트를 사용하여 구성되었기 때문에 원칙적으로 말뭉치 구축의 과정에서 임의로 추가된 문장은 존재하지 않는다. 따라서 경우에 따라 ‘있/았’의 대응이 영어에서 나타나지 않는 경우도 충분히 나타날 수 있는데, 이 경우에는 ‘없음’으로 태깅되었다. 예시는 아래와 같다.

- (14) 아니/VCN+오/EF+./SP 전혀/MAG 없/VA+였/EP+습니다/EF+./SF
 --- 없음
 No/UH ./, not/RB a/DT bit/NN ./.

둘째, 대응하는 영어 표현이 부정형(infinite)이거나 분사(particle)인 경우도 존재한다. 이 경우에는 과잉분석이 이루어지지 않도록 ‘무시제’로 태깅하였다. 구체적인 예시는 아래와 같다.

- (15) a. 37/SN 길/NNG+을/JKO 잃/V~~A~~+~~았~~/EP+을/ETM 때/NNG
 --- 무시제
 37/CD When/WRB Lost/VBN in/IN a/DT City/NN
- b. 글썸요/IC+./SF 한/MM 10/SN+분/NNB 정도/NNG
 기다리/VV+ㅏ ~~ㅏ~~/EP+다가/EC
 걸/VV+어/EC+보/VX+시/EP+지요/EF+./SF
 --- 무시제
 Well/UH ./, better/RBR wait/VB 10/CD minutes/NNS or/CC so/RB
 and/CC then/RB give/VB it/PRP a/DT try/NN ./.

(15a)의 경우에는 대응하는 영어 표현에서 시제를 담당하는 주동사가 생략된 구문이고, (15b)는 대응하는 문장이 명령법으로 구성되어 있다. 이와 같은 경우에는 시제 정보가 명시적으로 표현되지 않은 부정절인 까닭에, ‘무시제’로 주석처리하였다. 셋째, 원자료의 태깅 오류로 인해 한국어 문장과 영어 문장이 정확히 대응되지 않은 항목이 발견되었을 경우, ‘오류’로 태깅하였다. 이밖에도 한국어와 영어의 시제 및 상 관련 대응에서 특이 구문이 몇 차례 발견되었는데, 예시는 아래와 같다.

- (16) 어디/NP 가/VV+ㅏ ~~ㅏ~~/EP+다/EF 오/VV+ㅏ ~~ㅏ~~/EP+어/EF+?/SF
 --- 현재완료 / 없음
 Where/WRB have/VBP you/PRP been/VBN ?/.

(16)에서는 한국어의 서술어가 ‘갔다왔다’로 실현되어 있는데, 이 경우 ‘~~았~~/~~았~~’이 한 어절 내에서 연속적으로 출현하지 않았다는 점에서 ‘~~았~~’ 형태와는 차이가 있다. 대응하는 영어 표현에서 서술어는 단일어이기 때문에, 대응관계에서는 서술어의 대응쌍이 일대일이 아니게 된다.

끝으로 주석처리 과정에서 발견된 예외적 현상을 언급하면서 방법론에 대한 논의를 마무리하고자 한다. 세종 한영 병렬 말뭉치를 살펴보는 가운데 “한국어의 ‘~~았~~/~~았~~’은 과거시제를 표현하는 ‘필요조건’이다”라는 가정에 대한 반례가 발견되었는데, 회상표지 ‘더’가 사용되는 경우이다.

- (17) a. 남편/NNG+은/JX 금반지/NNG+를/JKO 주머니/NNG+에서/JKB
 꺼내/VV+**더**균요/EF+./SF
 My/PRP\$ husband/NN took/VBD out/RB a/DT gold/NN ring/NN
 from/IN his/PRP\$ pocket/NN ./
- b. 남편/NNG+은/JX 북/NNP+에/JKB 아내/NNG+와/JC
 자식/NNG+들/XSN+이/JKS 있/VA+다고/EC
 말/NNG+하/XSV+**더**균요/EF+./SF
 He/PRP told/VBD me/PRP that/IN he/PRP had/VBD a/DT wife/NN
 and/CC children/NNS in/IN the/DT North/NNP ./
- c. 한국/NNP+은/JX 여러/MM 면/NNG+에서/JKB 미국/NNP+과/JKB
 비슷/XR+하/XSA+**더**균요/EF+./SF
 I/PRP find/VBP Korea/NNP is/VBZ like/IN America/NNP in/IN many/JJ
 respects/NNS ./
- d. 잠꼬대/NNG+를/JKO 하/VV+ㄴ/ETM 때/NNG+도/JX 영어/NNP
 단어/NNG+를/JKO 외우/VV+**더**라구/EF+./SF
 Even/RB when/WRB he/PRP talks/VBZ in/IN his/PRP\$ sleep/NN
 /, it/PRP sounds/VBZ like/IN he/PRP 's/VBZ memorizing/VBG
 his/PRP\$ vocabulary/NN list/NN ./

(17a)와 (17b)의 경우, ‘있/았’이 출현하지 않았으나 대응하는 영어문장의 시제는 과거이며 직관적으로 판단하였을 때도 해당 한국어 문장은 과거에 일어난 사태를 기술하고 있다. 그러나 (17c)에서 예시된 바와 같이 ‘더’가 출현하였다고 하더라도 반드시 과거시제가 되는 것은 아니다. 또한 (17d)에서 보이는 것처럼, ‘더’는 때로 습관성을 나타내는 용법으로 쓰이기도 하며, 이때 시제 또한 영어에서는 과거가 아닌 현재이다. ‘있/았’의 기본 기능이 과거시제를 표현하는 것이라면, ‘더’는 기본적으로 회상(retrospective)의 서법 의미를 담당하는 표지로 파악된다(장석진 1995; 2012, Chang 2005, Sohn 2013). 그러나 회상이라는 것은 필수적으로 과거의 사태를 함의하는 것이기 때문에 그 논리구조 자체가 일정부분 과거시제 의미를 수반하고 있는 것으로 분석된다. 본 연구에서는 ‘더’를 분석의 대상에 포함시키지 않았으며, ‘더’의 시제 및 서법 관련 의미에 대한 말뭉치 분석은 추후 연구과제로 미루고자 한다.

이상에서 제시된 방법론에 입각하여 10,000여개의 ‘있/았’ 표지에 대한 의미 주석을 완료하였다. 주석처리 된 결과는 이어지는 4절과 5절에서 계량적 분석이 이루어진다. 4절에서는 주로 비율 중심의 기술통계 방법으로 한국어의 ‘있/았’의 쓰임을 밝힌다. 즉, ‘있/았’이 어떠한 시제와 상으로 대응되는가에 대한 전체적인 양상 분포를 계산하는 것이다. 다음으로 5절에서는 추론통계의 방법을 사용하여, ‘있/았’과 결합하였을 때 과거시제를 표현하지 않을 가능성이 높은 서술어의 목록을 정리하고 이들 서술어의 특성을 살펴보도록 한다. 이후

의 과정에서 텍스트의 처리는 파이썬(ver. 2.7)으로 구현되었으며, 모든 통계 연산과 차트 생성은 R(R Core Team, 2014)로 수행되었음을 밝힌다.

4. 비율 분석

전체 태깅된 ‘있/았’의 분포는 [표1]에 제시된 바와 같다. 시제를 ‘과거’, ‘현재’, ‘미래’로 구분하였을 때, ‘있/았’이 과거시제 의미를 담당하는 경우가 완료상과 진행상으로 실현된 항목을 포함하여 약 66%를 차지하고 있다. 이 가운데 정확히 ‘과거’의 의미로만 대응되는 경우가 60%를 상회한다. 이는 시제와 상과 관련하여 ‘있/았’이 수행하는 기능은 다양할 수 있으나, 그 대응이 ‘과거’일 가능성이 가장 크다는 것을 의미한다.

표 1 주석처리 결과

시제	의미	빈도	비율	비율합계
과거	과거	6580	61.94%	65.77%
	과거완료	240	2.26%	
	과거완료진행	3	0.03%	
	과거진행	164	1.54%	
현재	현재	863	8.12%	22.40%
	현재완료	1369	12.89%	
	현재완료진행	72	0.68%	
	현재진행	76	0.72%	
미래	미래	25	0.24%	0.28%
	미래완료	3	0.03%	
	미래진행	2	0.02%	
기타	무시제	111	1.04%	11.54%
	없음	1104	10.39%	
	오류	11	0.10%	

[표1]에서 ‘기타’항목으로 분류된 11.75%를 제외한 이후 비율을 다시 측정하면 [그림1]과 같다. ‘있/았’은 전체의 70% 이상 ‘과거’로 대응되고 ‘과거완료’, ‘과거진행’, ‘과거완료진행’을 포함하면, 전체의 75%가량이 ‘과거’와 관련을 맺고 있다.

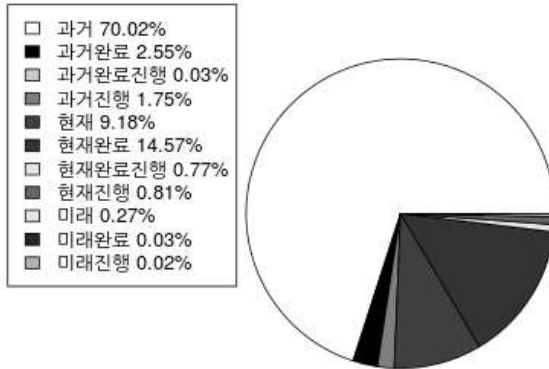


그림 1 각 의미대응관계의 비율분포

이는 ‘있/았’의 기본적 의미 기능이 과거시제의 표현이라는 주장이 실제 자료를 통해서 입증되었음을 의미한다. 그러나 75%가 100%는 아니다. 25%가량의 항목의 경우 다른 시제가 ‘있/았’ 표지에 할당된다. 구체적인 예시를 살펴보자.

가장 먼저 비율은 그다지 크지 않지만, 가장 흥미로운 경우로서 ‘있/았’이 ‘미래’에 대응되는 용례를 살펴보자.

- (18) a. 지금/NNG+쯤/XSN+은/JX 거기/NP+에/JKB
 도착/NNG+하/XSV+시/EP+ ㅏ /EP+을/ETM 것/NNB
 같/VA+습니까/EF+?/SF
 --- 미래
 Do/VBP you/PRP think/VB he/PRP 'll/MD be/VB there/RB
 by/IN now/RB?/.
- b. 그/NP+는/JX 멀리/MAG 가/VV+지/EC 앓/VX+았/EP+어요/EF+./SF
 --- 미래
 He/PRP won't/MD get/VB far/RB ./.
- c. 잠깐/MAG 쉬/VV+였/EP+다/EC 하/VV+ ㅊ 가요/EF+?/SF
 --- 미래
 Shall/MD we/PRP take/VB a/DT break/NN for/IN a/DT while/NN
 now/RB ?/.
- d. 맑아지/VV+ ㅓ /EP+으면/EC 좋/VA+겠/EP+어요/EF+./SF
 --- 미래
 I/PRP hope/VBP it/PRP 'll/MD clear/VB up/RP ./.
- e. 기다리/VV+ ㅓ /EP+다가/EC 나중/NNG+에/JKB 쓰/VV+도록/EC
 하/VV+죠/EF+./SF

--- 미래

I/PRP 'll/MD wait/VB and/CC use/VB the/DT phone/NN later/RB ./.

(18a-b)의 경우에는 대응하는 영어 표현이 조동사 will을 취하고 있다는 점에서 ‘미래’로 태깅되었으나 실제 이들 영어 문장에서 will은 ‘미래’의 의미라기보다는 ‘추측’의 의미에 가까울 수 있다. 영어에서는 MD로 태깅되는 조동사의 경우 ‘-ed’ 형태와 달리 중의적일 수 있다. 비슷한 맥락에서 (18c)의 shall 역시 미래 의미를 담당하는 조동사가 아닌 ‘청유’의 의미를 수행하는 조동사로 볼 수 있을 것이다. 실제로 최소한 현대미국영어에서의 shall은 미래조동사로서의 기능을 거의 상실한 것으로 알려져 있다. 그렇다면슨 치더라도 (18c)에서 “쉬는 행위” 자체는 미래에 일어나는 일이다. 흥미로운 점은 이때 한국어에서 ‘쉬/VV’가 과거시제 표지 ‘있/았’과 결합을 하였고, (18c)는 “잠깐 쉬다가 하실래요?”라는 표현만큼 자연스럽다는 것이다. (18d-e)의 예시도 마찬가지이다. (18d)에서 “날씨가 맑아지는 일”은 미래에 발생할 일이고, (18e)에서 “기다리는 행위”도 앞으로 발생할 일이다. 이 경우에 ‘있/았’의 사용은 형태와 의미의 불일치에 해당한다.

다음으로 전체의 9.19%를 차지하는 ‘현재’의 경우, 예시는 아래와 같다.

- (19) a. 잘되/VV+ㅏ ✖/EP+균요/EF+./SF
 --- 현재
 That/DT 's/VBZ good/JJ news/NNS ./.
- b. 아/IC+./SP 아주/MAG 예쁘/VA+게/EC
 생기/VV+ㅏ ✖/EP+균요/EF+./SF
 --- 현재
 Oh/UH ./, my/PRP\$!/. She/PRP 's/VBZ really/RB pretty/JJ ./.
- c. 319/SN 공항/NNG+까지/JKB 아직/MAG
 멀/VA+였/EP+습니까/EF+?/SF
 --- 현재
 319/CD Is/VBZ the/DT airport/NN much/RB further/RBR away/RB ?/.
- d. 거의/MAG 다/MAG 오/VV+ㅏ ✖/EP+균요/EF+./SF
 --- 현재
 we/PRP 're/VBP almost/RB there/RB ./.
- e. 백/NNP 선생/NNG+님/XSN+./SP
 전화/NNG+오/VV+ㅏ ✖/EP+는데요/EF+./SF
 --- 현재
 Mr./NNP Park/NNP ./, the/DT phone/NN 's/VBZ for/IN you/PRP ./.

(19a)와 (19b)에서 사용된 서술어는 ‘잘되/VV’와 ‘생기/VV’이다. 앞서 언급한

바와 같이 ‘있/았’을 필수적으로 요구하는 서술어가 별도로 존재하는데, 이들이 그 집합에 속한다. (19c)의 ‘떨/VA’의 경우에도 비슷하지만, “공항까지 아직 됩니까?”의 최소 대립쌍 역시 정문이라는 점에서 (19a) 및 (19b)와 다른 양상을 보인다. 다만 직관적으로 판단하였을 때, (19a-c)에서 사용된 ‘있/았’은 일반적인 의미의 과거시제와는 차이가 있는 듯하다. (19d)와 (19e)에서 사용된 ‘오/VV(또는 오/VX)’의 경우에는 특징적으로 ‘있/았’의 과거시제 의미에 부합하지 않는 특성을 보인다. 흥미로운 점은 ‘가/VV(또는 가/VX)’는 이러한 양상을 보이지 않는다는 점이다. 이는 ‘가다’와 ‘오다’라는 지시적 이동 동사(deictic verb)가 한국어와 영어가 가지는 의미적 차이에서 비롯된 것으로 추측한다.

끝으로 ‘현재완료’ 또는 ‘현재완료진행’에 대응되는 경우(14.57%, 0.77%)를 살펴보도록 하자.

- (20) a. 우리/NP+~ /JX 지금/MAG 몇/MM 시간/NNG+째/XSN
계속/MAG 이/MM 일/NNG+을/JKO
하/VV+ㄷ /EC+오/VX+ㄷ ✖/EP+습니다/EF+./SF
--- 현재완료
We/PRP 've/VBP been/VBN at/IN this/DT for/IN hours/NNS ./.
- b. 죄송/XR+하/XSA+지만/EC 아직/MAG 안/MAG
들어오/VV+시/EP+ㄷ ✖/EP+는데요/EF+./SF
--- 현재완료
I/PRP 'm/VBP sorry/JJ ./, but/CC he/PRP hasn't/VBZ returned/VBN ./.
- c. 어떻게/MAG 지내/VV+시/EP+ㄷ ✖/EP+습니까/EF+?/SF
--- 현재완료
How/WRB have/VBP you/PRP been/VBN ?/.
- d. 잊/VV+지/EC 않/VX+았/EP+습니다/EF+./SF
--- 현재완료
I/PRP haven't/VBP forgotten/VBN ./.
- e. 그동안/MAG 뵈/VV+고/EC 싶/VX+였/EP+어요/EF+./SF
--- 현재완료진행
I/PRP 've/VBP been/VBN looking/VBG forward/RP to/TO seeing/VBG
you/PRP ./.

(20a-b)에서는 위 (19d-e)와 마찬가지로 ‘오/VX’ 또는 ‘오/VV’가 사용되었는데, 과거시제의 일반적인 의미와 관련하여 ‘오/VX’와 ‘오/VV’는 사뭇 다른 구조를 취하고 있음을 엿볼 수 있다. 다음으로 (20c-e)에서 사용된 서술어(‘지내/VV’, ‘잊/VV’, ‘싶/VX’)의 경우에는 동사가 상적 의미를 어느 정도 포함하고 있는 요소로 판단한다. 즉, 이들 서술어가 ‘있/았’과 결합을 이루면 과거로부터 현재까지 지속적으로 이루어져 온 사태를 칭하는 의미 속성을 취하게 되는

것으로 분석한다.

이상의 예시에서 추론하건대, ‘있/았’이 과거시제 표지 이외의 기능을 수행할 때는 ‘있/았’과 결합하는 서술어의 의미 속성이 중요한 역할을 수행한다. 이를 바탕으로 각각의 서술어가 ‘있/았’과 결합하였을 때 과거시제와 어느 정도의 합치점을 보이는가에 대한 비율을 살펴보기로 하자. 어떠한 ‘서술어+있/았’의 형태가 가지는 의미적 기능의 전체 집합이 ‘과거’만을 요소로 가지는 경우, 여러 다른 의미를 가질 수 있으나 ‘과거’가 가장 큰 분포를 가지는 경우, 복수의 의미에서 ‘과거’가 가장 큰 분포가 아닌 경우, 그리고 ‘과거’로의 대응이 전혀 없는 경우로 나누어 비율을 다시 측정하면, 그 결과는 [그림2]와 같다. 이 과정에서 출현 빈도가 1회뿐인 서술어는 희박하게 출현하는 자료가 전체 양상을 파악하는 데 지장을 주는 요인이 될 수 있어 추출 과정에서 제외하였다.

[그림2]를 살펴보면 ‘있/았’과 공기하였을 때 전체 서술어 가운데 84% 이상이 더 빈번한 확률로 ‘과거’ 의미를 보이는 것으로 파악된다. 이는 ‘있/았’의 기본 기능이 과거시제 표현이라는 것을 다시금 증명하는 수치이다. 이외는 별도로 [그림2]에서 살펴야 할 사항은 ‘있/았’이 과거시제와 직접적 대응을 보이지 않는 요소들이다. 전체 서술어의 15%가량이 이러한 성향을 보이는데 구체적인 서술어 목록은 (21) 및 (22)에 제시되어 있다.

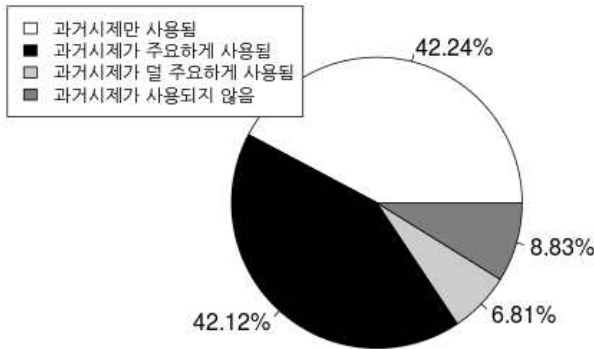


그림 2 서술어와 ‘과거’의 대응관계 비율

(21) 과거시제로 주요하게 대응되지 않는 서술어: 끝나/VV, 떨어지/VV, 바쁘 /VA, 변하/VV, 생기/VV, 오/VX, 이루/VV, 이룩/XR+하/XSV, 전화/NNG+드리/VV, 정하/VV+ㄱ/EC+지/VX, 지내/VV, 해내/VV, ...

(22) 과거시제로 대응이 이루어지지 않는 서술어: 닦/VV, 달라지/VV, 되 /VV+어/EC+주/VX, 발전/NNG+하/XSV, 살아오/VV, 실망스럽/VV, 정상화/NNG+되/XSV, 하/VV+ㄱ/EC+오/VX, ...

위 (21-22)에서 ‘끝나/VV’, ‘이루/VV’, ‘이룩/XR+하/XSV’, ‘해내/VV’, ‘달라지/VV’, ‘발전/NGG+하/XSV’, ‘하/VV+ㄷ/EC+오/VX’ 등과 같은 요소는 어휘상이 비교적 분명하게 드러나는 항목이다. 또한 ‘생기/VV’, ‘닭/VV’, ‘잘되/VV’ 등의 동사는 이미 기존 연구 등에서 ‘있/았’과 흥미로운 대응관계를 보인다는 점이 지적된 바 있다.

이러한 목록은 그 자체로 ‘있/았’의 쓰임에 대한 시사점을 제공해 주고 있으나, 보다 통계적으로 적합한 방법을 통해 목록을 정비할 필요가 있다. 예컨대, 위에서 제시된 ‘정하/VV+ㄷ/EC+지/VX’ 또는 ‘정상화/NGG+되/XSV’와 같은 항목은 출현 빈도가 비교적 낮다. 전자는 총 3회 출현하여 그 가운데 1회만 ‘과거’에 대응하고 있으며, 후자는 총 2회 출현하고 모두 ‘현재완료’에 대응된다. 이러한 대응은 혹 우연에 의해 이루어진 것일 수 있기 때문에 추가적인 검증을 통해 세밀한 해석을 수행해야 할 것이다. 다음 절에서 각 서술어들이 ‘과거’와 어떠한 대응관계에 보이는가에 관한 추론통계가 진행된다.

5. 상관 분석

각각의 서술어가 ‘있/았’과 결합하였을 때 과거시제와 얼마나 합치되는 확률분포를 보이는가를 확인하기 위해 본고가 사용하는 방법론은 Collexeme 분석이다(Stefanowitch & Gries, 2003, Gries & Stefanowitch 2004a, 2004b, Gries 2012). Collexeme 분석은 어떠한 어휘가 특정 언어적 속성과 어느 정도의 연관성을 가지는가를 직접적 확률 계산을 통해 검증하는 방법이다. 주어진 자료 내에서 어떠한 어휘 L에 대해서 L이 어떠한 구조적 속성 C와 공기하는 기능성은 네 가지로 나뉠 수 있다. 따라서 Collexeme 분석은 보통 2X2 표를 통해 도식화되는데, 아래의 [표2]를 보도록 하자.¹²⁾

12) 2x2 표를 사용하는 통계분석 방법으로는 피셔의 정확성 검증 이외에도 카이스퀘어 검증과 log-likelihood 검증이 존재한다. 일반적으로 피셔의 정확성 검증은 표본의 크기가 작을 때 사용하는 것으로 알려져 있으나, 필자가 이해하는 범위에서 더 정확히 설명하자면 피셔의 정확성 검증은 “표본의 크기에 따른 영향을 받지 않는 직접적 확률 계산”이다. 그럼에도 불구하고 피셔의 정확성 검증이 다른 두 2x2표 기반 검증에 비해 언어 연구에서 덜 빈번히 사용되는 이유는 아마도 직접적 확률 계산에 따른 연산량이 매우 크기 때문일 것이다. 피셔의 정확성 검증은 R의 기본 패키지로 포함되어 있어 R상에서는 손쉬운 연산이 가능하므로, 본 연구에서는 확률값의 직접적 비교를 위해 피셔의 정확성 검증을 사용하였다.

표 2 Collexeme 분석

	어휘 L	-어휘 L
구조 C	빈도(L ∩ C)	빈도(¬L ∩ C)
¬구조 C	빈도(L ∩ ¬C)	빈도(¬L ∩ ¬C)

각 서술어의 빈도 정보를 위 [표2]와 같은 2X2 표에 옮기고 나면, 이 수치들을 이용하여 Fisher의 정확성 검증을 시도한다(Fisher 1992). 예컨대, ‘보/VV’의 경우 주석처리된 자료에서 ‘있/았’과 총 96회 공기하여 출현하고 있는데, 이 가운데 ‘과거’로 태깅된 경우가 총 68회이다. 이를 기준으로 아래 [표3]을 생성할 수 있다.

표 3 보/VV

	보/VV	¬보/VV
과거	68	5522
¬과거	28	2284

이 각각의 수치를 사용하여 R에서 정확성 검증을 하면 아래와 같은 결과를 얻을 수 있다. 이때 영가설 H0은 “해당 서술어는 ‘있/았’과 결합하였을 때 과거시제와 상관이 없다.”에 해당할 것이고 대립가설 H1은 “해당 서술어는 ‘있/았’과 결합하였을 때 과거시제와 유의미한 상관관계를 지닌다.”이다.

- ```
(23) Fisher's Exact Test for Count Data
data: rbind(c(68, 5522), c(28, 2284))
p-value = 1
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
 0.6361585 1.6250756
sample estimates:
odds ratio
 1.004501
```

결과의 해석에 있어서는 유의수준 95%를 사용하였으므로, p값이 0.05보다 작아야 영가설을 기각할 수 있다. 다시 말해, 정확성 검증을 시도한 결과의 p값이 0.05보다 작으면 해당 서술어는 ‘있/았’과 결합하였을 때 “통계적으로 유의미하게 과거의 의미로 사용되는 경향이 선호되거나 또는 비과거의 의미로 사용되는 경향이 선호된다”고 해석할 수 있다. 위에서는 p값이 1이므로 ‘보/VV’



는 영가설을 기각하지 못한다. 즉, 보/VV'와 '있/았'의 결합은 다른 서술어에 비해 과거의 의미 또는 비과거의 의미 중 어떤 것에 대한 선호도를 가지고 있지 않다.

마찬가지의 방법을 사용하여 위 (21)에서 제시된 '생기/VV'와 '정하/VV+ㄷ/EC+지/VX'를 각기 검증해 보기로 하자. 두 서술어의 분포표는 각기 [표4] 및 [표5]와 같다.

표 4 생기/VV

|     | L  | -L   |
|-----|----|------|
| 과거  | 6  | 5584 |
| -과거 | 23 | 2289 |

표 5 정하/VV+ㄷ/EC+지/VX'

|     | L | -L   |
|-----|---|------|
| 과거  | 1 | 5589 |
| -과거 | 2 | 2310 |

이들 각각의 p값은 각기 3.2953268927654e-08, 0.206690238867252으로 측정 되었으므로 '생기/VV'는 영가설을 기각하는 반면 '정하/VV+ㄷ/EC+지/VX'는 영가설을 기각하지 못한다. 이는 유의수준 95% 범위에서 '생기/VV'는 비과거에 대한 선호도가 특징적으로 높다는 추론을 확정할 수 있는 반면, '정하/VV+ㄷ/EC+지/VX'는 비과거에 대한 선호도가 우연에 의한 것인지 또는 서술어 자체가 가지는 속성인 것인지를 확정할 수 없다는 뜻이다.

이러한 방식을 활용하여 세종 한영 병렬 말뭉치에서 '있/았'과 결합하는 모든 서술어 (빈도 1 제외)를 대상으로 Fisher의 정확성 검증을 실시하였다. 전체 계산은 (i) 분포표를 자동 산출하여 (ii) 이를 바탕으로 R스크립트를 자동 생성하고 (iii) 그 스크립트를 터미널상에서 일괄 처리하는 방식을 택하였다. 그 결과 Fisher의 정확성 검증을 통해 (iv) p값이 0.05 보다 작은 항목의 서술어를 별도로 추출하였다. 이상의 분석에서는 양측 검증을 활용하였기 때문에 '과거'와 유의미한 상관성을 지닌다는 점만 확인될 뿐, 그 상관성이 배타적인지 우호적인지는 확인되지 않는다. 어떠한 서술어가 '과거'의 의미와 배타적인가 아닌가를 확인하기 위해 위에서 추출된 서술어 목록을 대상으로 (v) 단측 검증을 추가로 실시하였다. 단측 검증은 결과에 대한 방향성을 지니기 때문에 검증의 두 대상이 선호의 관계에 있는지 배타의 관계에 있는지를 설명할 수 있기 때문이다. 즉, 단측 검증의 p값이 유의수준의 범위를 넘어선다면 이는 해당 서술어가 '있/았'과 결합을 하였을 때 비과거의 의미를 선호한다는 점을 입증한다. 이상의 과정을 통해 최종적으로 추출된 목록은 아래와 같다.

(24) 거두/VV, 걸/VV, 걸리/VV, 고장/NNG+나/XSV, 기여/NNG+하/XSV, 꿩/VV, 끝나/VV, 나/VV, 남/VV, 녀/VV, 다녀오/VV, 달라지/VV, 닭/VV, 되/VV, 되/VV+어/EC+주/VX, 떨어지/VV, 머무르/VV, 못하/VX, 믿/VV+어/EC+오/VX, 발전/NNG+하/XSV, 보/VX, 불/XPN+가능/NNG+하/XSA, 사라지/VV, 살/VV+아/EC+오/VX, 살아오/VV, 생기/VV, 서명/NNG+하/XSV, 시기/NNG+이/VCP, 실망스럽/VA, 안/MAG+되/VV, 안되/VA, 안정/NNG+되/XSV, 앓/VA, 예약/NNG+하/XSV, 오/VV, 오/VX, 이루/VV, 이룩/NNG+하/XSV, 이룩/XR+하/XSV, 읽/VV+어/EC+보/VX, 잘되/VV, 전달/NNG+되/XSV, 전화/NNG+드리/VV, 접어들/VV, 제공/NNG+하/XSV, 준비/NNG+되/XSV, 중/NNB+이/VCP, 지나/VV, 지내/VV, 질리/VV, 차/VV, 채택/NNG+하/XSV, 하/VV+키/EC+오/VX, 해결/NNG+되/XSV, 해내/VV, 화나/VV, 확인/NNG+되/XSV

(24)에서 제시된 서술어들은 과거시제 표지 ‘있/았’과 결합하였을 때 비과거가 높은 기대빈도로 선호되는 집단이다. 결과로 제시된 내용에서 특기할 만한 사항은 다음과 같다. 첫째, 피동형이 다수 포함되어 있다는 점이다. 이는 피동화의 과정에서 어휘의 상적 속성이 공고해 지기 때문이 아닌가 추측한다. 둘째, 어휘상이 비교적 뚜렷하게 파악되는 서술어가 다수 보인다. ‘끝나/VV’, ‘고장/NNG+나/XSV’, ‘이룩/NNG+하/XSV’, ‘이루/VV’, ‘사라지/VV’, ‘머무르/VV’ 등의 서술어가 여기에 속한다. 셋째, ‘닭/VV’, ‘생기/VV’, ‘잘되/VV’와 같은 서술어는 ‘있/았’과 결합을 하였을 때 특징적인 비과거 선호성을 보인다. 넷째, ‘오/VV’, ‘오/VX’ 및 관련 파생어가 다수 여럿 있는데 이는 ‘오다’와 ‘come’의 용법적 차이에 기인하는 것으로 본다. 끝으로 ‘전화/NNG+드리/VV’ 또는 ‘예약/NNG+하/XSV’와 같은 서술어가 포함된 것은 언어가 사용되는 문화적 맥락에 기인하는 것이 아닌가 생각한다.

## 6. 결론

본 연구는 한국어의 ‘있/았’의 분포와 관련하여 두 가지 항목을 검토하였다. 하나는 한국어의 ‘있/았’이 가지는 의미적 속성의 분포를 계산하여 과연 ‘있/았’의 기본적 기능이 과거시제의 표현이 맞는지를 확인하고자 하였다. 다른 하나는 한국어의 여러 서술어 가운데 어떠한 항목들이 ‘있/았’과 결합하였을 때 비과거 시제의 의미를 더 빈번하게 취하는가를 파악하고자 하였다.

이러한 연구 목적을 위해 본 연구에서는 세종 한영 병렬 말뭉치를 사용하였다. 한국어에서 ‘있/았’이 사용된 표현이 대응되는 영어 문장에서 어떻게 실현되는가를 기본 준거로 하여, 한국어 ‘있/았’이 가지는 용법에 대한 일관된

주석처리를 시도하였다. 그 결과 10,000개 이상의 항목에 대한 태깅이 이루어졌으며, 그 결과 약 75%의 비율로 ‘있/았’은 과거시제로 대응되고 있음이 확인되었다. 이는 ‘있/았’의 기본적 속성이 과거시제 표지임을 밝히는 수치이다. 그러나 ‘있/았’이 ‘현재’와 ‘현재완료’ 대응되는 비율 역시 각기 9%, 15% 정도로 추정되었다. 다른 한편으로 각각의 서술어를 대상으로 하여 어떠한 서술어가 더 빈번하거나 덜 빈번하게 과거시제로 대응되는가를 추론하였는데, 이 통계적 추론은 Fisher의 정확성 검증을 이용한 Collexme 분석으로 수행되었다. 그 결과 ‘있/았’과 관련하여 비과거 선호성을 유의미한 확률로 보이는 57개의 서술어를 추출하였으며, 어휘상을 비롯한 서술어의 의미 속성이 ‘있/았’의 의미 표상에 영향을 주고 있음을 확인하였다.

본고의 연구 결과는 한국어 ‘있/았’의 용법에 대한 전체적인 그림을 보여주고 있으나, 각각의 서술어에 대한 구체적인 용례분석은 추후 연구에서 보다 세밀하게 관찰되고 분석되어야 할 것이다. 또한 3절의 말미에서 언급한 바와 같이 ‘있/았’과 함께 때로 ‘더’가 과거시제에 대응하는 예시도 관찰되었는데, 이 역시 중요한 후행 연구 주제에 해당한다. 다른 연구자들이 본 자료를 활용하고 검토할 수 있도록 본 연구 과정에서 정리되고 계산된 결과 일체는 아래의 공간에서 다운로드할 수 있음을 밝힌다.

(25) <http://corpus.mireene.com/download/ed.html>

## 참고문헌

- 강남옥. 2011. “한국어 학습자의 시제 개념 습득 연구.” 「국어교육연구」 49. 239-268.
- 김문형, 조유미, 유현조, 장하연, 남승호, 신효필, 김윤신. 2012. “한국어 집합 시간 표현의 의미 유형과 표상.” 「언어와 정보」 16(1). 25-43.
- 김종복. 2004. 「한국어 구구조문법」. 한국문화사.
- 김창구. 2012. “한국어 과거시제 형태소 습득에 있어서 처리교수의 효과.” 「동북아 문화연구」 31. 5-25.
- 김현식. 2004. “한국어와 스페인어의 시제의미구조.” 「이중언어학」 25. 63-92.
- 남기심. 1978. “國語文法の 時制問題에 關한 研究.” 탑출판사.
- 문숙영. 2005. “한국어 시제 범주 연구.” 서울대학교 박사학위논문.
- 문숙영. 2008. “시제 어미 및 시제 상당 표현의 사용과 관련한 몇 문제.” 「한국어의 미학」 27. 45-73.
- 박동열. 2003. “한국어 시제와 법체계에 대한 연구.” 「프랑스어문교육」 16. 197-222.
- 박영준. 1998. “형태소 ‘-었-’의 통시적 변천.” 「한국어학」 8. 67-88.
- 박진호. 2011. “시제, 상, 양태.” 「國語學」 60. 289-322.
- 박현아, 오은정. 2014. “한국어의 시상 형식 해석에 관한 실험 언어학적 연구.” 「한국언어학회 2014년 여름학술대회 발표논문집」 113-119.
- 서상규 외. 2007. “21세기 세종계획 국어 특수자료 구축 연구보고서.” 국립국어원.
- 아소요코. 2004. “한국어 문법형태소 ‘-었-’의 의미에 관한 연구.” 서울여자대학교 석사학위논문.
- 양용준. 2002. “영어와 한국어의 상 표지 비교.” 「언어과학연구」 22. 141-158.
- 유현조, 장하연, 조유미, 김윤신, 남승호, 신효필. 2011. “한국어 TimeML : 텍스트의 사건 및 시간 정보 연구.” 「언어와 정보」 15(1). 31-62.
- 이효상. 1995. “다각적 시각을 통한 국어의 시상체계 분석.” 「언어」 20(3). 207-250.
- 장기남. 2013. “한국어 학습자의 시제 표지 사용 양상 - 종결형에서 {-었-} 사용 중심으로.” 「인문연구」 67. 1-28.
- 장석진. 1995. 「정보기반 한국어 문법」. 한신문화사.
- 장석진. 2012. “話行의 統合文法的 記述.” 「학술원논문집 (인문 · 사회과학편)」 51(1). 37-87.
- 황원미. 2003. “프랑스어 과거시제의 한국어 번역에 관한 연구 -소설 『이방인』을 중심으로-.” 「불어불문학연구」 56(2). 929-957.

- Chang, Suk-Jin. 2005. "Retrospective Mood in Korean: a Constraint-based Approach." In Saliko S. Mufwene (ed.), *Polymorphous Linguistics: Jim McCawley's Legacy*, 295-326. Cambridge, MA: MIT Press.
- Fisher, Ronald A. 1922. "On the Interpretation of  $\chi^2$  from Contingency Tables, and the Calculation of P". *Journal of the Royal Statistical Society* 85 (1): 87 - 94.
- Gries, Stefan Th. and Anatol Stefanowitsch. 2004a. "Extending Collostructional Analysis: a Corpus-based Perspective on 'Alternations'." *International Journal of Corpus Linguistics* 9(1). 97-129.
- Gries, Stefan Th. and Anatol Stefanowitsch. 2004b. "Co-varying Collexemes in the into-Causative." In Michel Achard & Suzanne Kemmer (eds.), *Language, Culture, and mind*, 225-236. Stanford, CA: CSLI.
- Gries, Stefan Th. 2012. "Collostructions." In Peter Robinson (ed.), *The Routledge Encyclopedia of Second Language Acquisition*, 92-95. London & New York: Routledge, Taylor & Francis Group.
- Iatridou, Sabine. 2000. "The Grammatical Ingredients of Counterfactuality." *Linguistic Inquiry*. 31(2). 231-270.
- Lee, Kiyong. 2008. "Formal Semantics for Interpreting Temporal Annotation." In P. G. J. van Sterkenburg (ed.), *Unity and Diversity of Languages*, 97-108. Amsterdam & Philadelphia: John Benjamins Publishing Co.
- Lee, Iksop and S. Robert Ramsey. 2000. "The Korean Language." Albany: State University of New York Press.
- R Core Team. 2014. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Sohn, Ho-Min. 2013. *The Korean Language*. Seoul: Korea University Press.
- Song, Sanghoun. 2014. "On Subjunctives in Korean: Exploiting a Bilingual Corpus." *Language and Information*. 18(1). 1-32.
- Stefanowitsch, Anatol and Stefan Th. Gries. 2003 "Collostructions: Investigating the Interaction between Words and Constructions.." *International Journal of Corpus Linguistics* 8(2). 209-243.
- 趙世開, 沈家煊. 1984. "漢語「了」字跟英語相應的說法." 「語言研究」1. 012.

- 송상현 (Song, Sanghoun)
- 소속: 난양이공대학교
- 전자우편: sanghoun@ntu.edu.sg
  
- 논문투고일: 2015. 02. 04.
- 논문심사수정완료일: 2015. 02. 23.
- 논문게재확정일: 2015. 03. 16.